

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |   |   |
|---|---|---|
| <p>(51) International Patent Classification<sup>4</sup> :<br/><br/>C07H 21/04</p>   | <p>AI</p>   | <p>(11) International Publication Number: WO 88/ 03533<br/><br/>(43) International Publication Date: 19 May 1988 (19.05.88)</p> |
| <p>(21) International Application Number: PCT/US87/02822<br/>(22) International Filing Date: 29 October 1987 (29.10.87)<br/><br/>(31) Priority Application Number: 927,258<br/>(32) Priority Date: 4 November 1986 (04.11.86)<br/>(33) Priority Country: US<br/><br/>(71) Applicant: SYNTRO CORPORATION [US/US]; 10655 Sorrento Valley Road, San Diego, CA 92121 (US).<br/><br/>(72) Inventors: FERRARI, Franco, A. : 7545 High Avenue, La Jolla, CA 92037 (US). RICHARDSON, Charles : 2949 Renault Street, San Diego, CA 92122 (US). CHAMBERS, James : 3956 Georgia Street, San Diego, CA 92103 (US). CAUSEY, Stuart, C. : 308 La Amistista Road, Del Mar, CA 92014 (US). POLLOCK, Thomas, J. : 4354 Robbins Street, San Diego, CA 92122 (US).</p> | <p>(74) Agent: ROWLAND, Bertram, Jr. Leydig, Voit &amp; May-<br/>er, 350 Cambridge Avenue, Suite 200, Palo Alto, CA<br/>94306 (US).<br/><br/>(81) Designated States: AT (European patent), AU, BE (Eu-<br/>ropean patent), CH (European patent), DE (Euro-<br/>pean patent), DK, FI, FR (European patent), GB<br/>(European patent), IT (European patent), JP, LU<br/>(European patent), NL (European patent), NO, SE<br/>(European patent), SU.<br/><br/>Published<br/>With international search report.</p> |   |

(54) Title: CONSTRUCTION OF SYNTHETIC DNA AND ITS USE IN LARGE POLYPEPTIDE SYNTHESIS

(57) Abstract

Methods for the production of large polypeptides containing repeating sequences of amino acids utilizing biochemi-  
cal techniques, specifically DNA sequences coding for the expression of the large polypeptides. Systems utilizing exoge-  
nous transcriptional and translational regions to control the production of the large polypeptides are also provided.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                              |    |                                       |    |                          |
|----|------------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria                      | FR | France                                | ML | Mali                     |
| AU | Australia                    | GA | Gabon                                 | MR | Mauritania               |
| BB | Barnados                     | GB | United Kingdom                        | MW | Malawi                   |
| BE | Belgium                      | HU | Hungary                               | NL | Netherlands              |
| BG | Bulgaria                     | IT | Italy                                 | NO | Norway                   |
| BJ | Benin                        | JP | Japan                                 | RO | Romania                  |
| BR | Brazil                       | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CF | Central African Republic     | KR | Republic of Korea                     | SE | Sweden                   |
| CG | Congo                        | LI | Liechtenstein                         | SN | Senegal                  |
| CH | Switzerland                  | LK | Sri Lanka                             | SU | Soviet Union             |
| CM | Cameroon                     | LU | Luxembourg                            | TD | Chad                     |
| DE | Germany, Federal Republic of | MC | Monaco                                | TG | Togo                     |
| DK | Denmark                      | MG | Madagascar                            | US | United States of America |
| FI | Finland                      |    |                                       |    |                          |

- 1 -

## CONSTRUCTION OF SYNTHETIC DNA AND ITS USE IN LARGE POLYPEPTIDE SYNTHESIS

5

The government has certain rights in this invention as a result of support provided by the Department of the Navy for the work leading to the present invention.

10

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of Application Serial No. 927,256, filed November 4, 1986.

15

### INTRODUCTION

#### Technical Field

20

The field is related to the production of high-molecular-weight polymers, either nucleic acids or peptides that are the expression products of the nucleic acids, and is particularly related to the production of high-molecular-weight peptides containing repeating sequences by biochemical processes, the peptide finding use as structured materials.

25

#### Background

30

35

Recombinant DNA technology has been applied in the isolation of natural genes and the expression of these genes in a variety of host cells. Typically, this technology has had utility in producing biologically active polypeptides, such as interferons or peptide hormones, which were impractical to produce in useful amounts by other means. It was also possible to produce modified proteins by isolating natural genes and utilizing the techniques of site specific, in vitro mutagenesis to alter these genes and thereby change the polypeptides produced. Other polypeptides have been

created by combining sections of various native genes to produce new polypeptides that are chimeric molecules of the several naturally occurring molecules.

With the advent of efficient and automated  
5 methods for the chemical synthesis of DNA, it has become possible to synthesize entire genes and to modify such synthetic genes at will during the course of synthesis. However, these various technologies have been applied to the production of natural or modified  
10 versions of natural polypeptides. There have been very few attempts to use these technologies to create substantially new polypeptides. In nature, polypeptides have a wide range of chemical, physical and physiological characteristics. Nevertheless there are commercial applications for which known, naturally  
15 occurring polypeptides are not appropriate.

While biotechnology is versatile, usually it has been limited in its applications to naturally occurring products or modifications of naturally occurring  
20 molecules. One great strength of organic chemical synthesis, by contrast, has been the ability to transform inexpensive carbon materials to a wide variety of polymeric molecules, including naturally occurring molecules, but most importantly entirely new chemical  
25 structures, such as polypropylene and polyacrylates, which have defined and predicted chemical properties not associated with naturally occurring molecules.

Such materials, particularly high-molecular-weight polymers containing repeating sequences of amino  
30 acids, have proven difficult to produce by biochemical means. The genes necessary for producing large peptides containing repeating units of amino acids were unstable and often underwent intermolecular recombination causing deletions of repeating units in the  
35 gene. The development of a biotechnology which would produce polymeric molecules by biological processes

similar to those available by organic synthesis would significantly broaden the range of applications of biotechnology.

5 Brief Description of the Relevant Literature

The cloning of multiple lactose operators up to four in tandem is disclosed by Sadler et al., Gene, (1980) 3:279-300. Hybrid bacterial plasmids containing highly repeated satellite DNA is disclosed by Brutlag  
10 et al., Cell, (1977) 10:509-519. The synthesis of a poly(aspartyl-phenylalanine) in bacteria is disclosed by Doel et al., Nucleic Acids Research, (1980) 3:4575-4592. A method for enriching for proline content by cloning a plasmid which codes for the production of a  
15 proline polymer was disclosed by Kangas et al., Applied and Environmental Microbiology, (1982) 43:629-635. The biological limitations on the length of highly repetitive DNA sequences that may be stably maintained within plasmid replicons is discussed by Gupta et al. in  
20 Bio/Technology, p. 602-609, September 1983.

SUMMARY OF THE INVENTION

Methods and compositions are provided for the production of polypeptides having repetitive oligomeric  
25 units by expression of a synthetic structural gene. The individual units coding for an oligomeric peptide sequence are varied as to nucleotide sequence utilizing amino acid codon redundancy. Long nucleic acid sequences are built up by synthesizing nucleic acid  
30 oligomers which express a plurality of individual repetitive peptide units, and the oligomers are joined to provide a polynucleotide of the desired length. Expression systems are used which provide for the growth of the subject host to high density prior to significant  
35 expression of the polypeptide product, followed by induction of expression to provide high yields of the polypeptide product, which can be isolated from the

host cells. In one embodiment, a system is employed where the transcription initiation system of the synthetic gene is not recognized by the host RNA polymerase and a gene expressing a functional RNA polymerase under inducible regulation is included in the host.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Plasmid pSY701 structure.

Figure 2: Immunoblots of polypeptide products using antibody to (A) beta-lactamase or to (B) gly-ala peptide.

Figure 3: Construction flowchart for plasmid pG10/SlpI.

Figure 4: Immunoblots of polypeptide products (A) T7gp10/SlpI with anti-Slp Ab, (B) T7gp9/SlpI with anti-Slp Ab, or (C) staining with Coomassie blue.

Figure 5: Construction flowchart for plasmid pSY856.

Figure 6: Time course for accumulation of the kanamycin-resistance gene product with the T7 system.

Figure 7: Construction flowchart for plasmid pSY857.

Figure 8: Construction flowchart for plasmid pSY980.

Figure 9: (A) Amido black stain of gel containing the product of beta-galactosidase/SlpIII gene fusion; (B) immunoblot of same product with anti-Slp antibody.

Figure 10: Construction flowchart for plasmid pSY1280.

#### DESCRIPTION OF PREFERRED EMBODIMENTS

Novel polypeptides are provided which are polyoligomers of repeating, relatively short, amino acid sequence units. The oligomers may be linked by spacers of different amino acid sequence. The novel polypeptides therefore contain repetitive amino acid

sequences and are particularly useful as fibrous proteins, including elastomeric. The gene encoding the repeating-unit-containing peptides is produced to particularly avoid problems previously associated with genes containing multiple repeating units.

The genes of the subject invention comprise multimers of DNA sequences encoding the same amino acid sequence unit, where two or more different multimers encoding different amino acid units may be joined together to form a block copolymer. The individual units will have from 4 to 30 amino acids (12 to 120 nt), more usually 4 to 25 amino acids (12 to 75 nt), particularly 4 to 8 amino acids, usually having the same amino acid appear at least twice in the same unit, generally separated by at least one amino acid. The units of the multimer coding for the same amino acid sequence may involve two or more nucleotide sequences, relying on the codon redundancy to achieve the same amino acid sequence.

For the most part the DNA compositions of this invention may be depicted by the following formula:

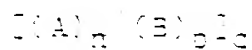
$$K_k (W M K_X N Y_Y)_i L_1$$

wherein:

K is a DNA sequence encoding an amino acid sequence of from about 1 to 100 amino acids, usually 1 to 60 amino acids, which may be any sequence, generally being fewer than about 20% of the total number of amino acids, more generally being fewer than about 10% of the total number of amino acids, which may be any sequence, particularly a naturally occurring sequence where the multimer structural gene has been fused to another DNA sequence in reading frame. K will have the initiation methionine codon.

k is 0 or 1;

W has the formula:



wherein:

5           A is a DNA sequence coding each time that it appears for the same amino acid sequence unit normally having at least one amino acid appear at least twice in the sequence, where A will generally be from about 12 to 90 nucleotides (nt), more usually for about 12 to 75  
10   nucleotides;

          where there will usually be at least two different A's, usually not more than ten different A's, more usually not more than six different A's, which code for the same amino acid sequence but differ from  
15   each other by at least one nucleotide and may differ by as many as ten nucleotides, usually not differing by more than about five nucleotides from another A sequence, each of the different A's usually being repeated at least twice; at least two different codons  
20   are employed for the same amino acid, e.g., GGC and GGA for glycine, in different A's coding for the same amino acid sequence unit;

          n will be an integer of at least 2, usually at least about 8, and not more than about 250, usually not  
25   more than about 200, frequently not more than about 125, and in some instances may not exceed about 50;

          B is a DNA sequence different from A coding for an amino acid sequence other than the amino acid sequence unit coded by the A unit and serves as a  
30   linking unit between oligomers of A units. B will generally have from about 3 to 45 nt, (1 to 15 amino acids) more usually from about 3 to 30 nt (1 to 10 amino acids);

          where the B units appearing in the gene may be  
35   the same or different, there usually not being more than about 10 different B units, more usually not more than about 5 different B units, where the B units may



differ in from 1 to 45 nt, more usually from about 1 to 15 nt, where the different B's may code for the same or different amino acid sequence;

p is 0 or 1 and may differ each time there is a successive A unit;

q is an integer of at least 1 and will vary with the number of nucleotides in A and B, as well as the values of n and p. The variable q will be selected so as to provide for at least 90 nucleotides for the multimeric portion of the structural gene, preferably at least about 150nt, more preferably at least 450nt, and most preferably at least 900 nucleotides, and the number of nucleotides will usually not exceed about 10,000, more usually not exceeding about 8,000, generally being in the range of about 900 to 6,000, more usually to about 5,000; and

M is a DNA nucleotide sequence of 0 to 18 nt, which may encode any amino acid sequence, usually including amino acids of A and/or B, generally limited to the amino acids of A and/or B;

X may be the same as or different from W, usually different, and will have the formula

$$[A^1]_{n^1} (B^1)_{p^1}]_{q^1}$$

wherein:

$A^1$ ,  $B^1$ ,  $n^1$ ,  $p^1$  and  $q^1$  are the same as or different from A, B, n, p and q respectively, at least one being different, wherein the analogous symbols come within the same definition as their counterparts;

x is 0 or 1;

N is the same as or different from M and comes within the same definition as M;

Y may be the same as or different from W, usually different, and will have the formula

$$[A^2]_{n^2} [B^2]_{p^2} [q^2]$$

wherein:

5  $A^2$ ,  $B^2$ ,  $n^2$ ,  $p^2$  and  $q^2$  are the same as or different from A, B, n, p and q respectively, at least one being different, wherein the analogous symbols come within the same definitions as their counterparts.

y is 0 or 1;

10 i is 1 to 100, usually 1 to 50, more usually 1 to 30, particularly 1, when x and y are 0;

when x or y are 1, q,  $q^1$  and  $q^2$  will be a total of at least 2, usually at least 5 and not more than about 50, usually not more than about 30.

15 The total number of nucleotides will be at least 90 nucleotides, usually at least about 150nt, preferably at least about 900nt and may be 20knt (kilo-nucleotides), usually not more than about 15nt, more usually not more than about 10knt.

20 The polypeptide encoded by the above DNA sequence will have the following formula:

$$K'_k (W' M' X'_x N' Y'_y)_i L'_1$$

wherein:

25 W' will have the following formula

$$[(D)_n (E)_p]_q$$

wherein:

30 D is the amino acid sequence encoded for by A and therefore has the numerical limitations based on 3 nucleotides defining a codon that codes for one amino acid;

35 E is the amino acid sequence encoded for by B, and therefore has the numerical limitations based on 3 nucleotides defining a codon, where each E may be the same or different, depending upon the coding of B;

and, wherein, likewise  $K'$ ,  $W'$ ,  $M'$ ,  $X'$ ,  $N'$ ,  $Y'$  and  $L'$  is the amino acid sequence encoded for by  $K$ ,  $W$ ,  $M$ ,  $X$ ,  $N$ ,  $Y$  and  $L$  respectively. However, in the case of  $K$  and  $L$ , subsequent processing, such as protease treatment, cyanogen bromide treatment, etc. may result in partial or complete removal of the N- or C-terminal non-multimeric chains.

$n$ ,  $p$ ,  $q$ ,  $k$ ,  $i$  and  $l$  have the same definitions as previously indicated.

Particular polymeric compositions having repeating multimeric units having the same compositions (A) will have the following formula where  $x$  and  $y$  are 0,

$$K'_k [(D)_n (E)_p]_q L'_l$$

where all of the symbols have been defined previously; and

the DNA sequence will have the formula

$$K_k [(A)_n (B)_p]_q L_l$$

where all of the symbols have been defined previously.

Particular DNA sequences encoding copolymeric compositions having a repeating unit of two to three multimeric blocks will have the following formula:

$$K_k (W'' M'' X'' N'' Y''_y)_i L_l$$

wherein:

$W''$  is a multimer having the formula

$$[(A^3)_{n^3} (B^3)_{p^3}]_{q^3},$$

where  $A^3$  is of 4 to 6, usually 4 to 6 codons, otherwise coming within the definition of A;

$n^3$  will be from about 2 to 12, usually 2 to 10;

$a^3$  is of from 2 to 8, usually 4 to 6 codons;

$p^3$  is 0 or 1;

$q^3$  is of from about 2 to 25, usually 2 to 20;

5  $X''$  and  $Y''$  are the same as or different from  $W''$ , usually different, coming within the same definitions as  $W''$ ;

$M''$  and  $N''$  come within the definitions of  $M'$  and  $N'$ ;

10  $i''$  is at least 2, usually at least 5 and not more than about 75, usually not more than about 50, generally not exceeding 30;

with the other symbols as defined previously.

The compositions of the invention will usually have a molecular weight of at least about 5kDal, usually 10kDal, preferably 15kDal and may have molecular weights as high or higher as 400kDal, usually not exceeding 300kDal, more usually not exceeding about 250kDal, the higher ranges generally being the multimer combinations, with the individual multimer usually being less than about 150kDal, usually less than about 100 kDal.

The nucleotide sequences which are employed will be synthesized, so that the repetitive units will have different codons for the same amino acid as described above. Usually, at least about 25%, more usually at least about 40%, and generally at least about 60%, but not greater than about 95%, preferably not greater than about 90% of the nucleotide sequences encoding the repetitive units will be the same.

30 Greater diversity within those ranges will be employed where the initial constructs are experimentally shown to undergo spontaneous recombination events.

Of particular interest are polypeptides which have as a repeating unit SGAGAG (G = glycine; A = alanine; S = serine). This repeating unit is found in a naturally occurring silk fibroin protein, which can be represented as GAGAG(SGAGAG)<sub>3</sub>SGAAGY (Y = tyrosine). In

the subject invention, the repeating unit is designed where the N-terminus may be MCAGAG or any other sequence of generally at least about 3 amino acids, usually at least about 5 amino acids, more usually 12 amino acids and not greater than 200, usually not greater than 100 amino acids, which may be different from the repetitive unit. Generally, a different N-terminus will be the result of insertion of the gene into a vector in a manner that results in expression of a fusion protein. Any protein which does not interfere with the desired properties of the product may provide the N-terminus. Particularly, endogenous host proteins, e.g. bacterial proteins, may be employed. The choice of protein may depend on the nature of the transcriptional initiation region. Similarly, the C-terminus may have an amino acid sequence different from the repeat sequence. Conveniently, there may be from 1 to 100, usually 1 to 25 amino acids, which may be the C-terminus of a naturally occurring structural gene, which again typically results from the formation of a fusion product.

A silk-like-protein (Slp) gene may be produced by providing oligomers of from about 5 to 25 repeat units as described above, more usually of about 10 to 20 repeat units. By having different cohesive ends, the oligomers may be concatemerized to provide for the polymer having 2 or more of the oligomeric units, usually not more than about 50 oligomeric units, more usually not more than about 30 oligomeric units, and frequently not more than about 25 oligomeric units.

The silk-like proteins may be varied by having alternate multimers with the same or different handedness. For example, in the formula,  $(B)_p$  may provide an even or odd number of amino acids. In silk, the hydrogens of the glycine may align on one side and the methyls and hydroxyls of alanine and serine on the other. If  $(B)_p$  is even, there will be continuous

alignment, if odd, there will be alternating alignment of  $(A)_n$ . Thus, different properties can be achieved by changing the number of amino acids encoded by  $(B)_p$ .

Of particular interest are polypeptides which  
5 mimic the composition and physical properties of silk of Bombyx mori.

Also of interest are polypeptides which have as a base repeating unit GVGVP (G = glycine, V = valine, P = proline), which may be found in naturally occurring elastin. In the subject invention, the  
10 N-terminus may be any convenient sequence and, if desired, may be in whole or in part removed by a protease. Usually the N-terminal sequence which does not have the subject motif will be less than about 100  
15 amino acids, more usually less than about 60 amino acids.

Of particular interest is a base sequence of about 6 to 10, preferably 8, units separated by a sequence of about 6 to 20 amino acids, usually 8 to 16  
20 amino acids, which may include an internal repeat different from the basic repeating unit of from 4 to 8 amino acids. For example, the second repeat sequence could be CAGACS, repeated twice. The total number of base repeating units will generally be in the range of about 150 to 300, usually 175 to 250. The C-terminus  
25 may terminate with a repetitive unit or portion thereof or a different sequence of from 1 to 100, usually 1 to 30 amino acids. The C-terminus is not critical to the invention and will be selected primarily for convenience. As with the N-terminus, it may be designed for  
30 proteolytic cleavage. As in the case of the silk protein, the subject elastin-like protein may be similarly engineered.

Of particular interest are proteins which  
35 mimic the properties of elastin and provide for elastomeric properties.

The copolymer involving repeating units is a powerful method for varying properties, by appropriate choice of the different units, the number of units in each multimer, the spacing between them, and the number of repeats of the multimer combination assembly. Thus, by varying the number and arrangement of primary monomers, a variety of different physical and chemical properties can be achieved.

Exemplary of the use of the block copolymers are combinations of silk units and elastin units to provide products having properties distinctive from polymers only having the same monomeric unit.

To prepare the structural genes, various approaches can be employed. To prepare the oligomers, complementary strands of DNA may be synthesized, so that upon hybridization double-stranded DNA is obtained with the appropriate termini. If desired, each of the oligomeric units may be the same, so that in a single step, a concatemer may be obtained depending upon the conditions of the hybridization. Normally, conventional annealing and ligating conditions will be employed, such as are described in the examples that follow.

If desired, two different oligomeric units may be prepared where the termini of the two units are complementary one with the other but the termini of the same unit are unable to bind together. In this way one can build individual oligomeric units and then join them together to form the concatemer, where the intervening linking sequences are defined at least in part by the termini. Depending upon the construct, the 5' terminus may provide for the initiation codon methionine, or the structural gene may be joined to an adapter which may provide for a unique sequence (optionally cleavable by a specific enzyme) at the 5' terminus or may be inserted into a portion of gene, usually endogenous to the host, in proper reading frame so as to

provide for a fusion product. By providing for appropriate complementary termini between the adapter or truncated gene and the 5' end of the subject structural gene, the sequences can be joined in proper reading  
5 frame to provide for the desired protein. Advantages that may be achieved by employing adapters or fusion proteins include having specific sequences for special purposes, such as linking, secretion, complex formation with other proteins, affinity purification, or the  
10 like.

Once the structural gene has been assembled, it may be cloned; clones having the desired gene, particularly as to sequence length, may be isolated; and the gene may be removed and used to join to a  
15 sequence for expression.

The expression construct will include transcriptional and translational initiation and termination regulatory regions, 5' and 3', respectively, of the structural gene. As already indicated, these regions  
20 may be created by employing a fusion protein, where the subject structural gene is inserted into a different structural gene downstream from its initiation codon and in reading frame with the initiation codon. Alternatively, various transcriptional and translational  
25 initiation regions are available from a wide variety of genes for use in expression hosts, so that these transcriptional and translational initiation regions may be joined to the subject structural gene to provide for transcription and translation initiation of the subject  
30 structural genes. A wide variety of termination regions are available which may be from the same gene as the transcriptional initiation region or from a different gene. Numerous constructs have been disclosed in the literature, and the same procedures may  
35 be applied with the subject gene as have been employed with other structural genes.



Of particular interest is the use of an inducible transcription initiation region. In this manner, the host strain may be grown to high density prior to significant expression of the desired product. Providing for inducible transcription is particularly useful where the peptide is retained in the cellular host rather than secreted by the host.

A number of inducible transcription initiation regions exist or can be employed in particular situations. The inducible regions may be controlled by a particular chemical, such as isopropyl thiogalactoside (IPTG) for inducing the beta-galactosidase gene. Other inducible regions include lambda left and right promoters; various amino acid polycistrons, e.g., histidine and tryptophan; temperature sensitive promoters; and regulatory genes, e.g., *cits* 857.

An alternative system which may be employed with advantage is use of a combination of transcription initiation regions. A first transcription initiation region which regulates the expression of the desired gene but which is not functional in the expression host by failing to be functional with the endogenous RNA polymerase is employed. A second transcription initiation region, such as an inducible region, can then be employed to regulate the expression of an RNA polymerase with which the first transcription initiation region is functional. In this manner expression only occurs upon activation of the regulatory region controlling the expression of the exogenous RNA polymerase. In the subject application, this system is illustrated with the T7 phage transcription initiation region, specifically the initiation regions of genes 9 and 10 of T7 phage.

An alternative system relies on the use of mutants which undergo a developmental change based on a change in the environment, such as a lack of a nutrient, temperature, osmotic pressure, salinity, or the

like. Illustrative of this system, strains of B. subtilis can be obtained which are incapable of sporulation but which can produce those components which initiate expression of products involved with sporulation. Therefore, by a change in the condition of the medium, a transcription initiation region associated with sporulation will be activated. In this situation, the host provides the necessary inducing agent or activator to initiate expression.

Various other techniques exist for providing for inducible regulation of transcription and translation of a gene in a particular host.

For the most part, the host will be a unicellular organism, either a prokaryote or a eukaryote, selected from bacteria, algae, fungi, filamentous fungi, etc. Illustrative hosts include E. coli, B. subtilis, B. stearothermophilus, S. cerevisiae, and the like.

The expression construct for expression of the desired gene, by itself or in conjunction with any auxiliary genes involved with transcription, will normally be joined to an appropriate vector for introduction into the expression host. A large number of vectors are commercially available with others being described in the literature. The vectors are normally characterized by having one or more unique restriction sites, a replication system for extrachromosomal maintenance in the host, and one or more markers which allow for selective pressure on the host. The markers may provide complementation, prototrophy to an auxotrophic host, resistance to a biocide, e.g., an antibiotic such as penicillin or kanamycin, or the like. In some instances, rather than selective pressure, a marker gene is employed which allows for detection of particular colonies containing the gene. This situation is illustrated by the gene for beta-galactosidase, where a substrate is employed which provides for a colored product.

The expression construct, including any auxiliary genes, may be introduced into the expression vector in accordance with known techniques, particularly employing restriction, insertion, and ligation.

5           The expression construct may then be used for transformation of the appropriate host. Depending upon the host, either intact cells or protoplast may be employed, where transformation or conjugation is employed. Conveniently, calcium-phosphate-precipitated  
10 DNA or non-ionic detergents may be employed to introduce the plasmid into the host. It should be appreciated that it is not necessary to employ vectors for host transformation, since bare DNA can be introduced for integration into the genome. However, even where  
15 integration is desired, a much greater efficiency of integration is achieved employing the vectors, thus favoring the employment of vectors.

          Depending upon the nature of the vector, the expression construct may be maintained on an extra-  
20 chromosomal element or become integrated into the host. Where integration is desired, it will usually be desirable with prokaryotes and some eukaryotes to have a sequence homologous to a sequence in the chromosome of the host. Usually the sequence will be at least  
25 about 200 bp and not more than about 5000 bp, usually not more than about 2000 bp. The choice of the homologous sequence is somewhat arbitrary, but may be used for complementation, where the host is an auxotrophic mutant and the homology provides prototrophy.

30           The transformants or integrants may then be grown in an appropriate nutrient medium to high density, followed by induction of transcription in accordance with the nature of the transcriptional system of the expression construct. Where the desired protein is  
35 retained in the cytoplasm, these cells are harvested and lysed, and, depending upon the use of the protein, the protein may be further purified in accordance with

conventional techniques, such as chromatography, solvent-solvent extraction, affinity chromatography, and the like.

The repetitive proteins can find a variety of uses. The Slp proteins may be used in producing fibers having unique properties, as a substitute for silk, and the like. Collagen proteins can be produced, where the collagen is free of the telopeptide or contains the telopeptide, depending upon its function. Atelopeptidecollagen should have little if any immunogenicity, so as to be a useful structural element for a variety of prosthetic devices or for use as a collagen substitute in other applications. Similarly, other proteins having repetitive sequences, such as keratin, can also be prepared in accordance with the subject invention. Other useful repetitive proteins can be prepared based on sequences of spider silks and other repetitive animal fibers. Artificial peptides useful for immunization could also be prepared based on repeating sequences present in various surface antigens of disease-causing microorganisms, such as parasites, bacteria, and viruses.

The following examples are offered by way of illustration and not with limitation.

#### Example 1

##### DNA Preparation Methods

##### 1. Preparation of plasmid DNA from E. coli:

A. Small scale: Plasmid DNA was prepared from 1.5 ml cultures by either the boiling procedure or the alkaline lysis method (Maniatis, et al., Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor. (1982)).

B. Large scale: A plasmid-carrying strain was grown overnight in 1 liter of Luria broth with the appropriate antibiotic. The cells were collected by

centrifugation at 10,000xg for 5 min and resuspended in 10 ml of ice cold TE (10mM Tris-HCl pH 8, 1mM EDTA). The cells were centrifuged again, resuspended in 1 ml of TES (TE and 25% w/v sucrose) and homogenized by vortexing. The samples were kept on ice for the following steps. Lysozyme (1 ml of 10 mg/ml) was added to the cell suspension and incubated for 5 min before the addition of 2 ml of 0.5M EDTA pH 8. After 10 min incubation, 50 ml of proteinase K (40 mg/ml) were added followed 10 min later with 15 ml of lysing buffer (0.1% Triton X-100, 1mM EDTA, 50mM tris-HCl pH 8). After 15-20 min, the cell lysate was centrifuged at 35,000xg for 90-120 min. The supernatant (19.8 ml) was transferred to a plastic tube with 20 g of CsCl and 400  $\mu$ l of ethidium bromide (10 mg/ml). After dissolution, the mixture was divided into two polyallomer ultracentrifuge tubes, sealed with heat and centrifuged in a Beckman Ti 65 motor at 60,000 rpm for 24 hr. The lower plasmid DNA band was removed from the tube with a hypodermic needle. The ethidium bromide was extracted three times with an equal volume of NaCl-saturated isopropanol. Two volumes of H<sub>2</sub>O were added to the DNA solution, and then the DNA was precipitated with ethanol.

## 2. Preparation of double-stranded DNA:

A culture of JM103 was grown to an OD<sub>500</sub> of about 0.2 and then divided into aliquots of 2 ml. Each aliquot was infected with a fresh plaque of M13 and incubated at 37°C for about 6 hr with vigorous shaking. Then the cells were pelleted and the supernatant was saved for subsequent infections. The double-stranded phage DNA was extracted by the boiling method (Maniatis et al.).

### 3. Deproteinization:

Phenol extraction was performed on a convenient volume of DNA sample, typically between 100  $\mu$ l to 10 ml. The DNA sample was diluted in 0.01M Tris-HCl pH 7.5, 1mM EDTA and an equal volume of water-saturated phenol was added. The sample was vortexed briefly and placed on ice for 3 min. After centrifugation for 3 min in a microfuge, the aqueous layer was removed to a new tube and extracted once with an equal volume of chloroform:isoamylalcohol (24:1).

### 4. Ethanol precipitation:

DNA in an aqueous buffer was concentrated by ethanol precipitation. To the DNA sample was added 1/10 volume of 3M sodium acetate pH 7.5 and 2-3 volumes of cold ethanol. The DNA was precipitated for 30 min at  $-70^{\circ}\text{C}$  or overnight at  $-20^{\circ}\text{C}$  and then pelleted by centrifugation in the microfuge for 15 min at  $4^{\circ}\text{C}$ . The pellet was washed once with 200  $\mu$ l of cold 80% ethanol and pelleted again for 10 min at  $4^{\circ}\text{C}$ . After air drying or lyophilization, the pellets were resuspended in the appropriate buffer.

### 5. Phosphatase treatment of DNA:

Phosphatase treatment of DNA was performed by adding 1  $\mu$ l (25 units) of calf intestinal phosphatase (Boeringer Mannheim) directly to the restriction enzyme digestion reaction and continuing the incubation for 30 min at  $37^{\circ}\text{C}$ . The phosphatase was inactivated for 60 min at  $65^{\circ}\text{C}$  prior to deproteinization by phenol extraction.

### 6. Fill-in reaction with DNA polymerase I:

DNA was resuspended in buffer containing 50mM Tris-HCl pH 7.4, 50mM KCl, 5mM  $\text{MgCl}_2$ , and 400  $\mu$ M each of the four deoxynucleotide triphosphates. Ten units of Klenow DNA polymerase (BRL) were added, and the

reaction was allowed to proceed for 15 min at room temperature. The DNA was then phenol extracted and ethanol precipitated.

5 7. T4 polynucleotide kinase reaction:

The reaction (10  $\mu$ l) contained: T4 polynucleotide kinase (BRL), 150 ng of DNA, 1  $\mu$ l of 10x kinase buffer (0.7M Tris-HCl pH 7.6, 0.1M  $MgCl_2$ , 50mM DTT) and [ $^{32}P$ ]-ATP (200-300 nCi). This was incubated  
10 at 37°C for 30 min and then the DNA was purified using a NACS column (Bethesda Research Labs).

8. Digestion with restriction endonucleases:

DNA was digested with restriction endonucleases (REN) in 1 x "AA" buffer [10 x AA buffer is 330 mM Tris-acetate, pH 7.9, 660mM potassium acetate, 100mM magnesium acetate, 50mM dithiothreitol (DTT) and 1 mg/ml bovine serum albumin (nuclease free)]. Whenever possible, the concentration of DNA was kept below 1  
15  $\mu$ g/25  $\mu$ l. Incubation was at 37°C for 1-4 hrs for most restriction endonucleases except for BalI, BanI and NaeI digestions which were incubated overnight.  
20

9. Analytical agarose gel electrophoresis of DNA:

To DNA samples for gel analysis we added 0.2  
25 volumes of loading buffer (5 x electrophoresis buffer, 0.01% bromphenol blue dye, 50mM EDTA, and 50% glycerol). Then the samples were loaded into lanes of a horizontal submerged electrophoresis unit containing  
30 a 1.0% (w/v) agarose gel. The electrophoresis buffer was either 1 x TAC or 1/2 x TBE. The 1 x TAC is 40mM Tris-base, 10mM EDTA, adjusted to pH 7.8 with acetic acid. The 1/2 x TBE is 0.045M Tris-base, 0.045M boric acid, 1mM EDTA, pH 8. The gel was run at 40-50V for 18  
35 hr, then removed and stained with 0.5  $\mu$ g/ml ethidium bromide for 30 min. The DNA bands were visualized on a long wavelength UV transilluminator.

10. Preparative agarose gel electrophoresis:

The procedures and materials are the same as for the analytical agarose gel electrophoresis. The only difference is the use of low melting point agarose, ranging in concentration from 0.5 to 2.5% (w/v) depending on the size of the DNA fragment to be purified. DNA restriction fragments were excised from the LMP agarose gels after visualization with ethidium bromide.

11. NACS purification:

Gel fragments containing DNA were melted at 70°C for 5 min and diluted approximately 5 fold with TE1 (10mM Tris-HCl pH 7.5, 0.2M NaCl). The gel solution was applied to a NACS column (BRL). The column was washed with 5 ml of the same buffer. The bound DNA was eluted with 300 µl of either TE2 (10mM Tris-HCl pH 7.5, 1.0M NaCl) for DNA fragments smaller than 1000 bp or TE3 (10mM Tris-HCl pH 7.5, 2M NaCl) for larger fragments. The eluted DNA was concentrated by ethanol precipitation.

12. DNA ligation:

Reactions for ligating cohesive ends contained: 1 µg DNA, 1 x AA buffer (see step 8, above) 1mM ATP and 20 units of T4 DNA ligase (BRL) in a 20 µl final reaction volume. The ligation was allowed to proceed for 16-18 hr at 15°C or 1-2 hr at room temperature. For blunt-ended ligations the reactions contained 1 µg DNA, 25mM Tris-HCl pH 7.5, 5mM MgCl<sub>2</sub>, 5mM DTT, 0.25mM spermidine, 200mg BSA, 1mM hexamine cobalt chloride (HCC), 0.5mM ATP and 400 units T4 DNA ligase (NEB) in a 20 µl reaction volume. The ligation was allowed to proceed for 30 min to 1 hr at room temperature.



Bacterial Transformation Methods1. Preparation of transformation-competent E. coli cells:

5 A culture of 200 ml of sterile L broth was inoculated with a small loopful of E. coli cells. This was incubated with shaking at 37°C until the OD<sub>600</sub> was approximately 0.5. The culture was placed on ice for 10 min and centrifuged at 6,000xg for 10 min. The cell pellet was resuspended in 100 ml of ice-cold 0.1M  
10 MgCl<sub>2</sub>, kept on ice for 30-40 min and centrifuged again. The pellet was resuspended in 2 ml of ice-cold 100mM CaCl<sub>2</sub>, transferred to a sterile test tube and incubated on ice for 24 hr. The competent cells were then aliquoted and stored at -70°C.

15

2. Transformation of E. coli:

An aliquot of frozen competent cells were thawed on ice. To 50 µl of cells 0.1 to 1 µg of DNA was added and the mixture was incubated on ice for 30  
20 min. The tube was removed from ice and placed in a 42°C bath for 2 min. L broth (1 ml) was added and the transformation mix incubated with shaking at the desired temperature (usually 30°C or 37°C) for 2 hr. Then one-tenth of the transformation was plated on L  
25 broth plates containing the appropriate antibiotic and, when necessary, XGAL and IPTG were added.

3. DNA transformation of B. subtilis:

B. subtilis cells were grown to early stationary phase (change in Klett units of 45% in 15  
30 min.). Transformation followed established procedures (Anagnostopoulos et al., 1981) (ref. 8). Cells (0.45 ml) were incubated with 1-10 µg of DNA at 37°C for 30 min with shaking, and then plated on TBAB agar plates  
35 with an appropriate antibiotic.

#### 4. Isolation of plasmid DNA from *B. subtilis*:

Plasmid DNA from *B. subtilis* was obtained by a method similar to the alkaline-lysis method except that pelleted cells were resuspended in 8 ml of solution 1 (50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH 8.0), 10 mg/ml lysozyme) and incubated at room temperature for 30 min. Then 16 ml of solution 2 (0.2N NaOH, 1% (w/v) SDS) was added and incubated on ice for 10 min. Finally, 12 ml of 3M potassium acetate (pH 4.8) was added and incubated an additional 20 min on ice. The lysed cells were centrifuged 15 min at 15,000 rpm in a Sorval SS-34 rotor. The DNA was precipitated by adding an equal volume of isopropyl alcohol and centrifuged at 7,000 rpm. The pellet was resuspended in 5 ml of 10mM Tris-HCl (pH 7.5), 1mM EDTA (TE). The solution was phenol extracted once and chloroform extracted. DNA was precipitated with ethanol and resuspended in 3 ml of TE. The volume was adjusted to 5.2 ml by adding 4.2g CsCl, 400  $\mu$ l of ethidium bromide at 10 mg/ml and TE. The solution was transferred to a Beckman quick-seal polyallomer centrifuge tube and centrifuged at 45,000 rpm in a Beckman vti65 rotor for 18 hr.

#### 25 Antibody Production, Protein Chemistry and Electrophoresis of Proteins

##### 1. Preparation of antibody to artificially synthesized peptides:

Synthetic peptide of sequence (GAGAGS)<sub>3</sub>GAAGY was coupled to BSA using the glutaraldehyde procedure of Kagen and Glick (1979). The degree of coupling was monitored using trace amounts of radioactive iodinated synthetic peptide. Peptide conjugates at a concentration of 1mg/ml in complete Freund's adjuvant were used to immunize rabbits at day 0. Animals were re-injected with antigen in Freund's incomplete adjuvant at day 30 and bitered at day 60. Positive

sera was detected using a microtiter RIA using the synthetic peptide as antigen. Kagen and Glick (1979), in *Methods of Radioimmunoassay*, Jaffe and Berman (eds.), Academic Press, p 328.

5           A peptide of 53 amino acids corresponding to the SlpIII sequence was prepared on an Applied Biosystems peptide synthesizer. The yield of this material, which has a molecular weight of 3640 was approximately 0.5 grams. The peptide was coupled to bovine  
10   serum albumin. The material was sent to Antibodies, Inc. for preparation of antibodies in rabbits. Antisera was obtained that reacted with synthetic peptides of both the SlpI and SlpIII sequences. These antisera have been useful for the detection of fusion peptides  
15   containing gly-ala sequences.

          Following the procedure described above an antigen was synthesized having the formula  $(V-P-G-V-G)_3$ , which was coupled to keyhole limpet hemocyanin. Polyclonal antisera was then prepared as  
20   described above which bound to the ELP peptide.

## 2. Polyacrylamide gel electrophoresis of proteins:

          Approximately  $10^9$  E. coli cells from growing cultures were pelleted by centrifugation at 10,000xg  
25   for 5 min. The cell pellets were resuspended in 100 to 500  $\mu$ l of 2X sample buffer (100mM Tris-HCl pH 6.8, 4% SDS, 10% B-mercaptoethanol, 60% glycerol or sucrose) and sonicated for 30 sec using a Tekmar sonic disruptor. Samples were boiled for approximately 5 min  
30   and 20 to 100  $\mu$ l of the cell lysates were loaded on an SDS-polyacrylamide gel (7.5 to 16% w/v). The gels were prepared following the procedure of Laemmli (Nature, 227:680-685 (1970)). The proteins in the gels were stained with 2% Coomassie brilliant blue in 10%  
35   methanol, 7.5% acetic acid for 1 hr and destained in 10% methanol, 7.5% acetic acid overnight.

### 3. Immunoblotting of proteins in gels:

After protein electrophoresis, one of the flanking glass plates was removed from the polyacrylamide gel. The gel surface was wetted with transfer buffer (25mM Tris-HCl, 192mM glycine, 20% methanol). A piece of nitrocellulose paper (Sartorius, SM11307) was saturated with transfer buffer and laid on the gel. Air bubbles between the filter and the gel were removed. The gel and nitrocellulose filter were placed in the transfer unit as specified by manufacturer (Bio-Rad). Transfer was allowed to proceed at 200 mA for 3-4 hr. Then the nitrocellulose filter was removed and stained with Amido-Schwartz for 3 min (0.05% Amido black, 45% deionized H<sub>2</sub>O), 45% methanol, 10% acetic acid) and destained in H<sub>2</sub>O. The filter was incubated for at least 10 min at room temperature in "BLOTTO" (5% w/v nonfat dry milk, 50mM Tris-HCl pH 7.4, 0.9% w/v NaCl, 0.2% w/v sodium azide). The filter was placed in serum appropriately diluted (1:50 to 1:500) in 0.5X Blotto (2.5% nonfat dry milk, 50mM Tris-HCl pH 7.4, 0.9% NaCl, 0.2% sodium azide) and was gently agitated for approximately 16 hr at room temperature. The filter was washed for 1 hr with 5 changes of TSA (50mM Tris-HCl pH 7.4, 0.9% NaCl, 0.2% sodium azide). The blot was placed in 15ml of 0.5X BLOTTO solution containing  $1 \times 10^7$  cpm of the <sup>125</sup>I -protein A and gently agitated for 2 hr at room temperature. The filter was washed for 2 hr with a minimum of 7 changes of TSA, rinsed once with deionized H<sub>2</sub>O and air dried. The blot was covered with Saran wrap and autoradiographed.

### 4. Amino Acid Analysis:

Amino acid compositions are determined by the PTC derivitization procedure of Henrickson and Meredith (1984). Protein samples were hydrolysed with 5.7 N constant boiling HCl at 108°C for 24 hours in vacuo. After reaction with PITC, amino acid derivatives were

detected at 254 nm by HPLC reverse phase chromatography using a Hewlett Packard 1090 system and a Supelco C18 column (4.6 mm x 25 cm) with a linear gradient of 0-50% acetonitrile in 0.1M NH<sub>4</sub>OAc pH 6.78 as a mobile base.

- 5 Henrickson, R.L. and Meredith, S.C. (1984) Amino Analysis by Reverse Phase High Performance Liquid Chromatography. Anal. Biochem. 137:65-74.

5. Amino Acid Sequence Analysis:

- 10 The N-terminal amino acid sequence was determined by automated Edman degradation using an Applied Biosystems Model 470A gas phase protein sequenator. The PTH amino acid derivatives were analyzed by reverse phase HPLC using a Hewlett Packard  
15 1090 system and an Altex C18 column (2 mm x 25 cm) with a complex gradient buffer system.

6. Peptide Synthesis:

- 20 Synthetic peptides were prepared by solid phase synthesis on an Applied Biosystems Model 430A Peptide Synthesizer using the standard symmetric anhydride chemistry as provided by the manufacturer. The coupling yield at each step was determined by the quantitative ninhydrin procedure of Sarin et al.,  
25 (1981). The synthetic peptide was cleaved from the solid support and amino acid blocking groups were removed using anhydrous HF (Stewart and Young, 1984). Crude peptides were desalted by chromatography over Sephadex G-50. Sarin, V.K., Kent, S.B.H., Tam, J.P.  
30 and Merrifield, R.B. (1981). Anal. Biochem. 237:927-936. Stewart, J.M. and Young, J.D. (1984). Solid Phase Peptide Synthesis, Pierce Chemical Company, Rockford, IL. pp 85-89.

### Synthetic DNA Methods

#### 1. In vitro DNA synthesis:

The N,N-diisopropylphosphoramidites, controlled-pore glass columns and all synthesis reagents  
5 were obtained from Applied Biosystems, Foster City, California.

Synthetic oligonucleotides were prepared by the phosphite triester method with an Applied Biosystems Model 380A DNA synthesizer using a 10-fold  
10 excess of protected phosphoramidites and 1  $\mu$ mole of nucleotide bound to the synthesis support column. The chemistries used for synthesis are the standard protocols recommended for use with the synthesizer and have been described (Matteucci, et al., Journal Amer. Chem.  
15 Soc., 103:3185-3319 (1981)). Deprotection and cleavage of the oligomers from the solid support were performed according to standard procedures as described by McBride, et al., Tetrahedron Letters, 24:245-248  
(1983). The repetitive yield of the synthesis as measured by the optical density of the removed protecting  
20 group as recommended by Applied Biosystems (1984) was greater than 97.5%.

The crude oligonucleotide mixture was purified by preparative gel electrophoresis as described by the  
25 Applied Biosystems protocols of November 9, 1984 (User Bulletin No. 13). The acrylamide gel concentration varied from 10 to 20% depending upon the length of the oligomer. The purified oligomer was identified by UV shadowing, excised from the gel and extracted by the  
30 crush and soak procedure (Smith, Methods in Enzymology, 65:371-379 (1980)).

#### 2. Sequencing of DNA:

DNA sequences were determined by the following  
35 methods. Fragments containing the region of interest were cloned into the multiple cloning site of M13mp18 or M13mp19 (Maniatis et al., 1982, and Norrander et

al., 1983). Single-stranded DNA was prepared and sequenced by the primer extension method (Sanger et al., 1977 and Biggin et al., 1983) using 35S-deoxyadenosine 5'-(alpha-thio)-triphosphate (New England Nuclear) as label. In some cases, reverse transcriptase (Molecular Genetics) was used to extend the primer, using the dideoxy:deoxynucleosidetri-phosphate ratios utilized by Zagursky et al. (Gene Anal. Techn. (1985) 2:89-94). Deoxyadenosine triphosphate labeled with either <sup>32</sup>P or <sup>35</sup>S was used in these reactions. Compression artifacts which appeared in some G-C rich sequences were overcome by eliminating deoxyguanosine triphosphate from the G reaction, and using deoxyinosine triphosphate (P-L Biochemicals) at a final concentration of 37.5  $\mu$ M instead. In the other mixes, the concentration of dideoxyGTP in the G reaction was 0.5 mM. All sequences were run on 6 or 8% polyacrylamide gels containing 8 M urea (Sanger et al. 1978). Primers used for sequencing were purchased from P-L Biochemicals. Storage and analysis of data utilized software from both DNASTar and International Biotechnologies, Inc.

### 3. In vitro mutagenesis of cloned DNA:

Plasmid DNA (1  $\mu$ g) containing the sequence to be mutated was digested in two separate reactions. One reaction contained either one or two restriction endonucleases which cleave at sites immediately flanking the region of interest. In the second reaction, the DNA was digested with a restriction endonuclease which cleaves only once at a site distant from the sequence to be mutated. The DNA fragments generated in the first reaction were separated by agarose gel electrophoresis and the large fragment which lacks the sequence to be mutated was excised and purified. DNA from the second reaction, the large fragment of DNA from the first reaction, and a synthetic oligodeoxynucleotide of 20-30 bases in length containing the

mutant sequence were mixed in a molar ratio of 1:1:250. The mixture was denatured by heating at 100°C for 3 min in 25 to 100 µl of 100mM NaCl, 6.5mM Tris-HCl pH 7.5, 6mM MgCl<sub>2</sub>, and 1mM β-mercaptoethanol. The denatured  
5 mixture was reannealed by gradually lowering the temperature as follows: 37°C for 30 min, 4°C for 30 min, and 0°C for 10 min. The reaction was supplemented with 0.5mM deoxyribonucleotide triphosphates, 1mM ATP, 400 units of T4 DNA ligase and 5 units of E. coli DNA poly-  
10 merase large fragment and incubated at 15°C for 12-16 hr. The reaction mixture was then transformed into E. coli and antibiotic-resistant colonies were selected.

#### Fermentation Conditions

15 The fermentor is a 15L Chemap, 10 L working volume. The culture conditions are: temperature = 30°C, pH = 6.8; NaOH 2.5 M is used for pH regulation. The headspace pressure is below 0.1 bar. The dissolved oxygen is regulated at 50%. The air flow varies from  
20 0.5 L/min to 20 L/min. The agitation rate varies between 200 to 1500 rpm.

The fermentor is inoculated with a 10% (v/v) inoculum grown in medium A for 15 hours at 30°C under agitation.

25 Medium B was the fermentor medium. The starting volume was 5 L.

When the glucose concentration reached 1%, a concentrated solution (5x) of medium B was added to the fermentor in order to keep the glucose concentration  
30 approximately at 1%. When the culture reached an OD<sub>600</sub> of 60.0, the temperature was increased to 42°C for 10 min, then lowered to 39°C for 2.5 hours. The cells were then harvested by centrifugation and frozen at  
-70°C until processed.

35



Table 1Medium A: LB Medium

| <u>Constituent</u> | <u>g/L</u>         |
|--------------------|--------------------|
| NaCl               | 10                 |
| tryptone           | 10                 |
| yeast extract      | 5                  |
| kanamycin          | $5 \times 10^{-3}$ |

Medium B

| <u>Constituent</u>                   | <u>g/L</u>           |
|--------------------------------------|----------------------|
| NH <sub>4</sub> Cl                   | 4.5                  |
| KH <sub>2</sub> PO <sub>4</sub>      | 0.75                 |
| MgSO <sub>4</sub> ·7H <sub>2</sub> O | 0.13                 |
| K <sub>2</sub> SO <sub>4</sub>       | 0.09                 |
| CaCl <sub>2</sub>                    | $24 \times 10^{-3}$  |
| FeSO <sub>4</sub> ·7H <sub>2</sub> O | $7.6 \times 10^{-3}$ |
| TE                                   | 0.5 ml               |
| casamino acids                       | 25                   |
| yeast extract                        | 5                    |
| glucose                              | 20                   |
| kanamycin                            | $5 \times 10^{-3}$   |

25

30

35

Example 2Assembly and Expression of the SlpI Gene1. Summary of the scheme for assembling the SlpI gene:

An 18 bp DNA sequence that codes for the most frequent repeating oligopeptide in the silk fibroin protein made by Bombyx mori [Lucas, F. and K.M. Rudall (1986) Extracellular Fibrous Proteins: The Silks. p 475-558, in Comprehensive Biochemistry, vol. 26, part B., M. Florkin and F.H. Stotz (eds.) Elsevier, Amsterdam] was synthesized in vitro. Two single-strands were synthesized, annealed together and then the resulting double-stranded segments were multimerized head-to-tail to generate concatamers of up to and exceeding 13 repeats. The structural gene for silk I that we proceeded to work with had 13 repeats that coded for the oligopeptide gagags, where g = glycine, a = alanine and s = serine. We refer to this structural gene as the "monomer". We constructed "dimeric, trimeric, tetrameric, pentameric and hexameric" SlpI genes containing 26 (SlpI-2), 39 (SlpI-3), 52 (SlpI-4), 65 (SlpI-5) and 78 (SlpI-6) repeats. There is a short intervening sequence between each monomer unit. The assembly is pictured as follows:

Repeating DNA sequence 5'-G G T G C G G G C G C A G G A A G T  
C G C C C G C G T C C T T C A C C A-3'

"Monomer"



Multimers



2. Assembly of the "monomeric" SlpI structural gene:

The two single-strands shown above were synthesized as previously described. The strands were separately purified by gel electrophoresis, phosphorylated using T4 polynucleotide kinase and then mixed together and allowed to anneal. This resulted in the double-stranded segments aligning spontaneously head-to-tail in long concatamers. The phosphodiester bonds between segments were formed with T4 DNA ligase. The reaction was stopped by filling in the terminal cohesive ends using the Klenow fragment of DNA polymerase I. The blunt-ended repeating DNA was then ligated to the HincII REN site in plasmid vector pUC12 (Veiera, et al., Gene 19:269-263 (1982)). The ligated DNA was transformed into E. coli HB101 and transformants were selected for their ability to grow in the presence of ampicillin. The DNA of potential clones was analyzed; for size and orientation by REN digestion and gel electrophoresis. DNA sequences were determined for isolates with large inserts that were oriented properly. The "monomer" clone selected for subsequent multimerization had 13 repeats coding for the oligopeptide agagsg, and was named pSY708. The DNA sequence, deduced amino acid sequence and REN sites of the SlpI insert and flanking regions of pSY708 are shown in Table 2.

30

35

5  
 10  
 15  
 20  
 25  
 30  
 35

A A G C T T C G C G C T G C A G G T C A C C C C G C C C G C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T  
 T T C G A A C C C C A C C T C C A G T G G G C C C C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A

k l g l q v z r a s a s s s a s a s s s

G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C  
 C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C

a s a s s s a s a s s s a s a s s s a s

G C A G G A A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A  
 C G T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T

a s s s a s a s s s a s a s s s a s a s s s

A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T G C C G C C C C A G G A A G T G G T  
 T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A C G C C C C G C T G C C T T C A C C A

s s a s a s s s a s a s s s a s a s s s

X B A S E  
 B A Y M C  
 A M A A R  
 I I I I I  
 I I I I I

G C G G G C C C A G G A A G T G G A C T G T A G A G G A T C C C C G C C C A G C T C G A A T T C  
 C G C C C C G C T G C C T T C A C C C T G A G A T G T C C T A G G G G C C C C G C T C G A C C T T A A G

a s a s s s g t l o d p r a s s s

230

### 3. Construction of the expression vector, pSY701:

Plasmid pSP65 (10 µg, Boehringer Mannheim) was digested with AatII REN, phenol extracted and ethanol precipitated. The DNA was resuspended in 10µl of H<sub>2</sub>O. One-half of this DNA was digested with exonuclease III in the following mix: 5 µg DNA, 10 µl 10X exonuclease III buffer (600mM Tris-HCl pH 8.0, 6.6mM MgCl<sub>2</sub>, 10 mM β-mercaptoethanol) and 9 units of exonuclease III in a total volume of 200 µl. Samples of 20 µl were taken at 0, 1, 2.5, 5 and 7.5 min and diluted immediately in 100 µl of the following buffer (30 mM sodium acetate, pH 4.5, 0.25 M NaCl, 1mM ZnSO<sub>4</sub>) containing 5 µg tRNA and 36 units of S1 nuclease. Incubation was at 30°C for 45 min and then the reaction was terminated by the addition of 15 µl of stop buffer (0.5M Tris pH 9.0, 125mM EDTA, 1% w/v SDS, 200 µg/ml tRNA). The samples were phenol extracted and ethanol precipitated. The resuspended DNA was digested with SmaI REN and electrophoresed through a 1% gel of low melting point agarose. The gel band corresponding to the DNA fragment carrying the β-lactamase gene, the plasmid origin and the β-galactosidase promoter was excised from the gel and melted at 65°C. One volume of H<sub>2</sub>O was added. The DNA in each sample (timepoint) was recircularized by ligation in the presence of agarose. The reaction included 8 µl melted gel, 2 µl of ligation buffer (100mM Tris-HCl pH 7.5, 50mM MgCl<sub>2</sub>, 50mM DTT, 1mM ATP), 10 units T4 DNA ligase and was incubated at 15°C for 3 hr. Competent cells of JM101 were transformed with the ligated DNA and transformants were selected by growth on L broth plates containing ampicillin (40 µg/ml). Plasmid DNA was prepared from four transformants. The DNA was digested with BamHI REN, labeled with <sup>32</sup>P-dGTP using the Klenow fragment of DNA Polymerase I, digested with Pvu I and then the smallest fragment was gel purified. The fragment from one transformant was sequenced using the Maxam and Gilbert

technique. The fragments of the other three plasmids were further digested with TaqI and electrophoresed on the same gel. The sequenced plasmid had a fusion between the multiple cloning site and a position upstream from the N-terminal ATG of  $\beta$ -lactamase. The size of the BamHI-TaqI fragment of two of the other plasmids indicated a fusion between the multiple cloning site and the 4th amino acid of the  $\beta$ -lactamase gene. The DNA and corresponding amino acid sequences of the N-terminal region of the altered  $\beta$ -lactamase are given below, along with a circular map of REN sites for pSY701 (see Figure 1). The amino acid sequence of Figure 1 is met-thr-met-ile-thr-pro-ser-leu-gly-cys-arg-ser-thr-leu-glu-asp-pro-his-phe-arg-val-ala-leu-ile-pro-phe-phe-ala-ala-phe-cys-leu-pro-val-phe-ala-his.

4. Insertion of "monomer" SlpI from pSY708 into pSY701:

Plasmid pSY708 was digested with HindIII, the cohesive ends were filled in using the Klenow fragment of DNA polymerase I and then digested with BamHI. Plasmid pSY701 was digested with XbaI, filled in as above and then digested with BamHI. The DNA fragment from pSY708 and the backbone of pSY701 were then purified by electrophoresis through a low melting temperature agarose gel and purified with NACS (BRL) columns. The appropriate fragments were mixed, ligated, and then transformed into E. coli JM109. Transformed cells were selected by growth on L plates containing ampicillin (40 mg/ml), IPTG ( $5 \times 10^{-4}$  M) and XGAL (20 mg/ml). Transformants were analyzed for plasmid contents and one (pSY756) was selected for further study since it carried the insert of the monomer SlpI-1 sequences in the proper orientation, as determined by mapping of REN sites. Although the entire DNA sequence was not deter-

mined for pSY756, the junctions between the insert and vector were verified as correct restriction sequences for XbaI, upstream and BamHI, downstream.

5     5.     Multimerization of the SlpI gene of pSY756:

Plasmid pSY708 was digested with the REX SmaI and the DNA fragment carrying the coding sequence for the polypeptide arg(ala-gly-ala-gly-ser-gly)<sub>13</sub>thr-leu-glu-asp-pro (R(AGAGSG)<sub>13</sub>TLED<sub>P</sub>) was purified as in 4 above. Plasmid pSY756 was digested with SmaI, deproteinized and then ligated with the purified DNA fragment from pSY708. Transformants of E. coli JM109 were selected on medium containing ampicillin. Clones were found to contain 2 units (dimer pSY882), 3 units (trimer pSY883), and 4 units (tetramer pSY915) of the original monomer sequence of the pSY708 clone. Similarly, pentamers and hexamers have also been constructed. All of these plasmids are genetically stable and produce the gly-ala peptide as a fusion with 15     β-lactamase.

6.     Expression of the SlpI gene fusion to the β-lactamase protein:

Synthesis in E. coli cells of the SlpI peptide as a fusion protein with β-lactamase was detected by immunoblotting (Western) analysis. Anti-"Slp" antibodies were raised against a synthetic silk peptide. Fusions between β-lactamase and SlpI were also detected with antibodies raised against the E. coli β-lactamase. As shown in Figure 2, this antibody reacts with dimers and trimers of SlpI fused to the E. coli β-lactamase. The SlpI insert proceeds the fifth amino acid of the signal sequence for this enzyme. The β-lactamase antibody (Figure 2A) detects both the unprocessed fusion proteins as well as the processed mature enzyme which 35     appears as the major antigenic band in this figure, at

about the 28 kD position. The mobilities of all SlpI-containing polypeptides are anomalously slow and the proteins are not as large as they appear on the gels.

The anti-SlpI antibody also is useful in detecting these fusion products. Lanes 2-5 of Figure 23 represent 4 separate clones that contain dimer fusions of SlpI with  $\beta$ -lactamase, while lanes 6 and 7 are from two clones containing trimer fusions. As can be seen the antigenicity of the trimer is considerably greater than for the dimer. It is known from prior experiments that fusion proteins containing only a monomer of SlpI are not detected at all with this anti-SlpI antibody. The increased antigenicity of the trimer peptide allows it to be detected as a processed fusion with the  $\beta$ -lactamase signal peptide. The processed form is seen at about the 33 kD position in lanes 6 and 7 of Figure 23. The appearance of normally processed  $\beta$ -lactamase mature enzyme (detected with  $\beta$ -lactamase antibody) as well as a peptide corresponding to the fusion between the SlpI-3 trimer and the signal peptide of  $\beta$ -lactamase (detected with gly-ala antibody) suggests that despite the insertion of SlpI sequences within the signal sequence, normal proteolytic processing of the enzyme occurs in E. coli.

25

#### 7.a. Expression of the SlpI gene by fusion to T7 genes:

The SlpI sequence has also been expressed as a fusion protein with both the gene 9 and gene 10 proteins from bacteriophage T7 in E. coli. The construction is diagrammed in Figure 3. Plasmid pSY915 (containing the SlpI-4 tetramer) was digested to completion with REN SalI and partially with BamHI. The DNA fragment containing the SlpI-4 tetramer was purified and then cloned in plasmid pSY114 (pG2 of Promega Biotech) which had been digested with RENs SalI and BamHI. From this intermediate plasmid, the tetramer insert of SlpI was removed with the RENs AccI and EcoRI. This frag-

30

35



ment was then cloned in pSY633 (pBR322 containing the complete T7 gene 9 sequence; pAR441 of Studier *et al.*, (1986)) which was digested with EcoRI and AsuII. In the resulting plasmid the SlpI tetramer is fused to the gene 9 translational reading frame near the C-terminus of gene 9. This plasmid was then used to transform E. coli strain 0-48 (strain HMS174 ( $\lambda$ DE3) of Studier, *et al.*, 1986) which contains the T7 RNA polymerase gene inserted into the chromosome under transcriptional control of the IPTG-inducible  $\beta$ -galactosidase promoter. In this configuration, expression of the SlpI-4 sequence is dependent upon production of the T7 RNA polymerase which itself is controlled by the IPTG-inducible  $\beta$ -galactosidase promoter. As shown in Figure 4B and 4C, when these cells are induced with IPTG a protein product of the gene 9/SlpI-4 fusion gene is synthesized and is detected with antibody to the synthetic Slp peptide. The fusion product migrates in the gel as if it was 82 kD in size. The size expected is only 65 kD. The anomalous mobility is characteristic of the unusual amino acid composition (rich in glycine and alanine) and is seen for all Slp-containing products.

In like manner, plasmid pSY638 (pAR2113 of Studier) containing the promoter region and the first 13 amino acids of the T7 gene 10 protein, was digested with REN BamHI, filled in with the Klenow fragment of DNA polymerase and then digested with REN EcoRI. Into this linearized plasmid was cloned the AsuII-EcoRI fragment of pSY633, containing the SlpI-4 tetramer. This ligation creates an in-frame fusion of the silk tetramer following the thirteenth amino acid of T7 gene 10. The latter fusion product may be used for spinning without further processing since the N-terminal 13 amino acids are only a small part of the large SlpI protein. Although the fusion product is about 30 kD in size, it has an anomalous mobility and migrates as if it was larger, 50 kD. This is shown in Figure 4A.

The plasmids pG9/SlpI-4 and pG10/SlpI-4 were further improved by inserting a kanamycin-resistance gene in the  $\beta$ -lactamase gene in the orientation opposite to the T7 expression system. Thus, any low level expression from the T7 system does not lead to elevated  $\beta$ -lactamase activity. Such activity eliminated the ampicillin in the medium that was added to select for maintenance of the plasmid. When the ampicillin was depleted the plasmids were lost from the culture. The kanamycin-resistance gene circumvents this problem and represents a significant improvement in the T7 expression system, especially for large scale cultures. The kanamycin-resistance gene (originally from Tn903) was isolated from a plasmid pUC4K (Veira, J. and J. Messing, 1982. Gene. 19:259-263) was a HincII fragment.

#### 7.b. Fermentation and purification of SlpI-4:

E. coli strain 0-48 carrying pSY997 was grown at 37°C, using a Chemap or a Braun fermentor, in 10L of LB to an OD (Klett units) of 300 ( $3 \times 10^9$  cells/ml). The T7 system was then induced with the addition of 3.5 mM IPTG. After 150 min the cells were concentrated 10x using a Millipore filter unit (Pellicon cassette system, 100,000 molecular weight cut off filter). The cell suspension was then frozen at -70°C until processing.

The cell suspension was melted in a water bath at 42°C and lysed in a french press, and the lysate was spun at 125,000xg for 1 hour at 25°C. The cleared supernatant was treated with DNAase (250  $\mu$ m/ml) for 15 min at room temperature, then filtered through a .45  $\mu$ m sterile filter. The filtrate volume was measured and incubated in ice with slow stirring. Then 231 mg of ammonium sulphate were added for each ml of filtrate over a period of 45 min. One ml of NaOH for each 10 g of ammonium sulphate was added to neutralize the pH.

After 2 hours of continuous stirring the mixture was spun at 9,000kg for 10 min. The pellet was resuspended in 1/10 of the original filtrate volume using distilled water. The centrifugation and resuspension was repeated three times. The pellet was resuspended in 1/10 of the original filtrate volume in distilled water. Samples were analyzed for protein concentration, amino acid composition and protein sequence by standard methods. This is one of several methods for obtaining the product. This method results in an SlpI-4 product that is greater than 90% pure. The amino acid composition is almost entirely gly, ala and ser, as expected, and the N-terminal amino acid sequence is that of the gene 10 leader.

15

8. Controlled expression of the T7 RNA polymerase gene in Bacillus subtilis:

The coding sequence of the T7 RNA polymerase gene (T7 gene 1, T7 nucleotides 3128 to 5845) from plasmid pSY558 (pAR1151 of Studier, et al., 1986) was modified by in vitro mutagenesis of cloned DNA. We inserted the recognition sequence for the restriction endonuclease NdeI at position 3171. Using an oligodeoxynucleotide which was synthesized as previously described, the T7 gene 1 sequence was changed from its natural sequence, TAAATG, to the modified sequence, CATATG.

Similarly, the upstream regulatory sequence of the Bacillus subtilis gene spoVG, obtained from plasmid pCB1291 (Rosenblum, et al., J. Bacteriology, 148:341-351 (1981)), was modified by in vitro mutagenesis at position 85 (Johnson, et al., Nature, 302:800-804 (1983)) such that it also includes an NdeI cleavage site. The upstream regulatory sequences of the spoVG gene were then ligated with the coding sequence of the T7 RNA polymerase gene via these novel NdeI cleavage sites. After transformation of E. coli HB101, the

35

plasmid contents of individual ampicillin-resistant isolates were checked by restriction mapping. The correct construction was named pSY649.

Plasmid DNA containing the *spoVG*:T7 RNA polymerase fusion gene (pSY649) was further modified to include a chloramphenicol-resistance gene that functions in *B. subtilis*. First the *Nde*I to *Sal*I fragment of about 1200 base pairs from plasmid pGR71-P43 (Goldfarb, *et al.*, *Nature*, 293:309-311 (1981)) was isolated. This fragment carries the P43 promoter of *B. subtilis* and an adjacent chloramphenicol acetyltransferase gene from Tn9. After filling in all the cohesive ends using the Klenow DNA polymerase reaction, this fragment was inserted into the *Xba*I site within the multiple-cloning site of pUC13 (Veiera, *et al.*, *Gene*, 19:259-268 (1982)). Ampicillin and chloramphenicol-resistant transformants were selected for further use. The correct plasmid construction was named pSY630. The *Sma*I to *Hinc*II endonuclease cleavage fragment from plasmid pSY630 containing the chloramphenicol acetyltransferase gene fused to the P43 promoter sequence was gel purified and blunt-end ligated to the *Pvu*I site of plasmid pSY649 that had been treated first with T4 DNA polymerase. The resulting plasmid, pSY856, was then transformed into *B. subtilis* I168. Because plasmid pSY856 is unable to replicate autonomously in *B. subtilis*, stable transformants resistant to chloramphenicol must result from the integration of the plasmid into the *B. subtilis* chromosome (Ferrari, *et al.*, *J. Bacteriology*, 154:1513-1515 (1983)). The integration event, facilitated by homologous recombination, most likely occurred at either the *spoVG* or the P43 loci of the bacterial chromosome (pSY856 contains DNA sequences homologous to the *B. subtilis* chromosome at only these two sites). The resulting strain, "BIPoL", was grown both in the presence and absence of chloramphenicol in order to deter-

mine the stability of the selectable marker. Expression of the T7 polymerase was obtained and this has no apparent effect on the growth or viability of this strain.

5

9.a. Expression of a plasmid-borne target gene  
(kanamycin-resistance) in B. subtilis strain  
BIPoL:

The Staphylococcus aureus plasmid pUB110  
10 (Lacey, et al., J. Med. Microbiology, 7:235-297, 1974)  
which contains the gene coding for resistance to the  
antibiotic kanamycin was used to test the expression of  
the growth-regulated spoVG:T7 RNA polymerase gene of  
strain BIPoL. An EcoRI-BamHI fragment of phage T7 DNA  
15 (positions 21,402 to 22,858 containing the T7 gene 9  
promoter sequence was purified from plasmid pAR441  
(Studier, et al., 1986). This DNA fragment was ligated  
into pUB110 between the EcoRI and BamHI restriction  
endonuclease sites. The resulting plasmid, pSY952,  
20 contains the T7-specific promoter in the same orientation  
as the kanamycin-resistance gene. Plasmid  
pSY952 was transformed into B. subtilis I168 and BIPoL  
and these strains were analyzed for the level of expression  
of the polypeptide encoded by the plasmid-  
25 derived kanamycin-resistance gene. Approximately 10<sup>9</sup>  
cells from growing cultures of I168, I168 containing  
pUB110, I168 containing pSY952, BIPoL, BIPoL containing  
pUB110, and BIPoL containing pSY952 were obtained at  
several times during the growth and sporulation cycle.  
30 The proteins in these cell samples were processed and  
analyzed by polyacrylamide gel electrophoresis.

Because the rate of transcription from the  
spoVG promoter increases as a function of cell density  
and reaches a maximum during early sporulation, an  
35 accelerated accumulation of the target protein is expected  
in the BIPoL strain containing pSY952 during  
growth as the culture enters sporulation. The results

show that a protein of molecular weight 34 kilodaltons increases in abundance as the culture approaches and enters stationary phase. The size of the protein is in agreement with the predicted size of the kanamycin-resistance gene product (Sadaie, et al., J. Bacteriology, 141:1173-1182 (1980)) encoded in pSY952. This protein is not present in B1Pol or I168 containing pSY952 which lacks the spoVG-regulated T7 RNA polymerase gene or in B1Pol containing pUB110 which lacks the T7 promoter sequence. The maximum accumulated level of target protein after 24 hours of growth in B1Pol containing pSY952 was 20% of the total cellular protein as determined by densitometry.

9.b Expression of SlpI-4 in B. subtilis:

Plasmid pG10SlpI was digested with EcoRI REN. After filling in the cohesive ends using the Klenow DNA polymerase reaction, the DNA was digested with BglII REN. Plasmid pSY662 was digested with SmaI and BamHI RENs. The two plasmids were then purified by electrophoresis through a low melting temperature agarose gel and purified with NACS (BRL) columns. The DNA fragment of pG10SlpI was ligated to the backbone of pSY662 and transformed into E. coli containing ampicillin (40 ug/ml). Transformants were analyzed for plasmid contents and one (pSY662/G10/SlpI-4) was selected for further study.

Competent cells of B. subtilis B1Pol were transformed with pSY662/G10/SlpI-4 and incubated at 37°C with shaking for 90 min. The transformation mixture was then diluted 1:100 in fresh LB containing 10 ug/ml of tetracycline and incubated at 37°C with shaking. Samples were taken and equal numbers of cells were lysed and loaded on gels for separation by SDS-PAGE. Immunoblot analysis was performed using anti-Slp antibodies to detect the synthesis of the gene 10/SlpI-4 fusion protein.

The expression of the SlpI-4 polypeptide in B. subtilis was detected by its seroreactivity with anti-Slp antibody, after transfer of the cellular proteins from the polyacrylamide gel to a nitrocellulose filter. We verified that the seroreactive protein was the product of the SlpI-4 gene by exhaustively treating the cellular proteins with CNBr. This should cleave after methionine residues, but since SlpI-4 lacks methionine it will remain intact. The CNBr treatment eliminated greater than 98% of the proteins stainable with Coomassie blue dye. And as expected for a protein lacking methionine, SlpI-4 remained intact and still reacted with anti-Slp serum.

15

### Example 3

#### Assembly and Expression of the SlpIII Gene

##### 1. Summary of the scheme for assembling the SlpIII gene:

20 The synthetic SlpIII gene codes for a protein similar to the SlpI gene and to the crystalline region of the silk fibroin protein made by the silkworm, Bombyx mori. SlpIII more closely resembles the silk fibroin molecule because it includes the amino acid tyrosine at regular intervals (about 50 residues), whereas multimers of SlpI do not. The SlpIII gene was assembled from smaller parts. First, three double-stranded sections of DNA of about 60 bp in length were chemically synthesized. Each section was cloned by  
25 insertion into bacteriophage M13 and the DNA sequence was verified. These sections were then removed from the vector and linked together in a specific order. This linkage of about 180 bp is named the SlpIII "monomer". "Monomers" were then linked in a specific  
35 order to yield dimers, trimers, tetramers, etc., of SlpIII. The multimers were then cloned either directly into plasmid expression vectors to detect the SlpIII

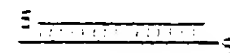
protein or initially into an adapter plasmid. Insertion of the SlpIII DNA into the adapter allows for further gene manipulation and is further described later. The assembly scheme is pictured as follows:

5

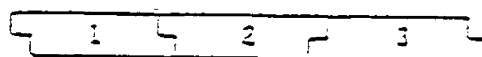
Synthesis of double-stranded DNA sections --

10 Section 1 - 

Section 2 - 

15 Section 3 - 

Assembly of "monomer" --



Multimerization --

20



The DNA and corresponding amino acid sequences of the three sections of the SlpIII gene are shown in Table 3.

30

35





The double-stranded DNA sequence is shown in the 5' to 3' direction. The amino acids (g = glycine, a = alanine, s = serine, y = tyrosine, coded by the sequence are shown immediately below each section.

5 Recognition sequences for cleavage by restriction endonucleases are shown above each section.

The above six single-strands were synthesized. After synthesis, the strands of DNA were purified and the homologous strands were annealed. About 1  $\mu$ l (0.5  $\mu$ g) of each strand was mixed with 2  $\mu$ l of 10X AA (description) buffer and 16  $\mu$ l of sterilized deionized H<sub>2</sub>O in a 1.5 ml polypropylene Eppendorf tube. The tube was placed in a boiling water bath (500 ml in a 1 liter beaker) for 10 min and then the beaker was removed from the hot plate and allowed to cool on the bench to room temperature. This required about 1-2 hr.

Each of the three double-stranded sections was cloned separately into M13mp18. Section 1 was ligated between the SmaI and BamHI restriction sites of the multiple-cloning site. Section 2 was ligated between the BamHI and PstI sites. And section 3 was inserted between the PstI and HindIII sites. The respective clones are: M13mp18.1, M13mp18.2, M13mp18.3. The DNA sequence was determined for each cloned section. One representative of each section that had the correct DNA sequence was recovered and became the material for the next step: assembly of the "monomer".

### 3. Assembly of the "monomer" of SlpIII:

30 The DNA sections 2 and 3 were isolated by digestion of the M13 clones with restriction enzymes: for section 2, M13mp18.2 was digested with BamHI and PstI; for section 3, M13mp18.3 was digested with PstI and HindIII. The two sections were purified and mixed together in equal molar amounts with M13mp18.1 that had been first digested with BamHI and HindIII. T<sub>4</sub> DNA ligase was added to link the homologous overlapping

ends in the order 1-2-3. Due to the hybridization specificity of the cohesive ends, the three sections are efficiently linked in only this order. The DNA sequence of the cloned "monomer" in the assembly named M13mp18.1.2.3 was determined to be correct and as shown in 2 above.

#### 4. Multimerization of the "monomer" of SlpIII:

In order to prepare large amounts of the "monomer" structural gene we first subcloned the "monomer" into the plasmid vector pUC12. M13mp18.1.2.3 was digested with EcoRI and HindIII restriction enzymes. The SlpIII "monomer" was gel purified and ligated into pUC12 digested with EcoRI and HindIII. The resulting plasmid DNA was prepared, the "monomer" was released from the vector by digestion with BanI REN and the fragment was gel purified.

To create multimers, "monomer" DNA with BanI ends were linked by ligation. The nonpalindromic terminal BanI recognition sequence allows linkage only in a head-to-tail order. The extent of multimerization is monitored by gel electrophoresis and staining the DNA with ethidium bromide. Multimers of more than 20 units have been obtained by this method.

#### 5. Cloning of the multimers of SlpIII:

Plasmid pCQV2 (Queen, et al., J. Appl. Mol. Gen., 2:1-10 (1983)) was digested with EcoRI and BamHI restriction endonucleases and a fragment of about 900 bp was purified. This DNA fragment contains the bacteriophage lambda cI-857 repressor gene, the closely linked rightward promoter, P<sub>R</sub>, and the beginning of the cro gene. Plasmid pSY335 (described as pJF751 in Ferrari, et al., J. Bacteriology, 161:556-562 (1985)) was digested with EcoRI and BamHI restriction enzymes and subsequently ligated to the DNA fragment of approximately 900 bp of pCQV2. The plasmid obtained from

this construction, pSY751, expresses the  $\beta$ -galactosidase gene at 37°C and 42°C, but not at 30°C (Figure 6).

In this approach the SlpIII gene is first  
5 cloned into an "adapter" sequence in an intermediate plasmid and then subcloned to the expression systems. The adapter sequence has the following useful  
features: a unique central BanI REN site, three unique  
REN sites to either side of BanI, information coding  
10 for protein cleavage at either methionine, aspartate-proline or arginine amino acids and small size. The BanI site is the point of insertion for the SlpIII multimers with BanI ends.

The adapter was synthesized with the Applied  
15 Biosystems 380A Synthesizer, cloned in M13mp18 and the DNA sequence verified. The adapter was then subcloned into a specially-constructed plasmid vector that lacked BanI REN sites. The recipient plasmid was made as follows. Plasmid pJH101 (Ferrari, et al., 1983) was  
20 partially digested with AhaIII restriction enzyme and religated. Transformants of E. coli HB101 were selected on medium containing chloramphenicol (12.5 mg/ml). After restriction analysis of several isolates one plasmid was chosen, pSY325 (Figure 7). This plas-  
25 mid contains only the chloramphenicol-resistance gene and the replication origin (from pBR322) of pJH101. After digestion to completion with XhoII, pSY325 was ligated with the gel-purified adapter. The result was the adapter-plasmid, pSY937, and its new REN sites were  
30 verified.

The SlpIII multimers were cloned into the BanI  
site of pSY937 (Figure 7). Positive clones were  
identified by colony hybridization and with the lower  
strand of section 1 of SlpIII as the DNA probe for  
35 hybridization (probe sequence shown in Table 2). Positive clones were characterized by gel electrophoresis for the size of the inserted multimer. Finally, the

SlpIII sequences were subcloned using the REN site in the flanking adapter regions to specific locations of expression plasmids.

The SlpIII protein had the following amino acid composition:

SlpIII            1178 AA            MW 83,000

(fm)    DPVVLQRRDWNPCVLTQLNRLAAHPPFASDPM  
 GAGS (GAGAGS)<sub>5</sub> GAAGY  
 [(GAGAGS)<sub>9</sub> GAAGY]<sub>18</sub>  
 GAGAGSCAGAGSCGAGAMDPCGRYQLSAGRYHYQLVWCQK

(fm) intends the initiation codon

#### SlpIII Expression Vector

Plasmid DNA pSY1086 is a pSY937 derivative containing 19 repeats of SlpIII (3.5 kb). This plasmid DNA was digested with NruI and PvuII and the fragments separated by agarose gel electrophoresis. The purified SlpIII multimer was then cloned in plasmid pSY751 digested with PvuIII REN. Several clones were analyzed and one (pSY1008) was chosen to be used in expression experiments and SlpIII purification.

The ampicillin drug resistance gene of pSY1008 was substituted with the Kanamycin marker from pSY1010 (produced by digestion of pSY633 with DraI and SspI and insertion of Kan<sup>R</sup> obtained by HincII digestion of pUC4K) and the subsequent plasmid was called pSY1186. By removing the SlpIII portion of plasmid pSY1186 with BanI, a new plasmid, pSY1262, was generated. This plasmid contains a unique BanI site which allows for the direct ligation of fragments containing BanI ends obtained by polymerization of monomers. This plasmid has been used to generate plasmids containing inserts for the following proteins: SELP1, 2, 3, and Slp4.

Production and Purification of SlpIIICell Culture

E. coli are cultured in the following medium:

|    |  |            |
|----|--|------------|
| 5  |  | <u>g/l</u> |
|    | yeast extract  | 20         |
|    | casamino acids                                       | 20         |
|    | peptone  | 20         |
|    | gelatin peptone                                      | 20         |
| 10 | KH <sub>2</sub> PO <sub>4</sub>                      | 2          |
|    | K <sub>2</sub> HPO <sub>4</sub>                      | 2          |
|    | Na <sub>2</sub> HPO <sub>4</sub> · 7H <sub>2</sub> O | 2          |
|    | glucose  | 2          |
|    | ampicillin   | 0.1        |

15

An overnight culture (500 ml - 1 L) which had been grown at 30°C was used to inoculate 375 L of media contained in a 500 L fermentor. Fermentor conditions include a tachometer reading of 100 rpm, vessel back pressure of 5 psi and an air flow of 170 l/min in order to maintain dissolved O<sub>2</sub> at greater than 50%.

Glucose (1 gm/l) and ampicillin (0.05 g/l) were added to the fermentation when the culture reached an OD<sub>650</sub> of 1.0 and again at 2.0. When the culture reached an OD<sub>650</sub> of 2.0 the temperature was increased to 42°C for 10 min and then lowered to 38°C for 2 hours. The culture was then chilled to 10°C and cells were harvested by centrifugation in a continuous centrifuge and frozen at -70°C until processed. Yields from two separate fermentations were 7.3 kg and 5.2 kg wet weight of cells.

It should be noted that other media can be used and, with different plasmids, various selection conditions can be imposed (i.e., substitution of kanamycin selection for ampicillin). These conditions have been used in laboratory scale fermentations (10 L volumes).

Cell Lysis

Method 1. Cells are thawed and suspended to a concentration of 1 kg wet weight/6 l in 50 mM Tris-HCl pH 7.0, 1 mM EDTA and broken by 2 passages through an APR Gaulin cell disrupter at 8000 psi. During this lysis procedure the cells are kept cold with an ice bath. The cell lysate is then centrifuged at 26,000xg with a continuous centrifuge, such as the T2-28 rotor in a Sorvall RC5B refrigerated centrifuge operated at 4°C. Under these conditions greater than 90% of the SlpIII produced can be found in the pellet. The supernate does contain some product which can be recovered by  $\text{NH}_4\text{SO}_4$  precipitation as described below. The pellet is extracted with LiBr as described below.

Method 2. Frozen cells are thawed and re-suspended to a concentration of 1 kg wet weight/6 l in 50 mM Tris-HCl pH 7.0, 10 mM EDTA, and 5 mM PMSF to inhibit protease activity. Cells are stirred in this buffer at room temperature for 0.5 to 2 hours, then lysozyme is added to a concentration of 1 g/l and incubation is continued for 20 min.  $\beta$ -Mercaptoethanol is then added to 70 mM and the detergent NP40 is then added to a final concentration of 1% for 20 min while continuously stirring the cell suspension. Then  $\text{MgCl}_2$  is added to 50 mM followed by DNase at a concentration of 1 mg/l and incubation is continued at room temperature for 20 min. The cell lysate is then centrifuged as in method 1 at 26,000xg in a continuous centrifuge and the supernatant is collected and passed through the continuous centrifuge a second time at 26,000xg. The supernate resulting from this second centrifugation contains <5% of the total SlpIII, but what is there can be recovered with  $\text{NH}_4\text{SO}_4$  as described below. The pellets resulting from the 1st and 2nd 26,000xg centrifugations are combined and extracted with LiBr as described below.

Method 3. For this method, a strain of E. coli is used that contains a second plasmid which encodes the T7 phage lysozyme. This plasmid is compatible with the plasmid encoding the SlpIII gene and the drug resistance determinant. The strain is grown in the same medium and under the same conditions as in the first two methods. However, due to the production of the T7 lysozyme inside the cells, their cell wall is weakened and they can be easily lysed at the completion of the fermentation by the addition of EDTA to >100 mM and NP40 to a concentration of from 0.5 to 1.0% v/v. Lysis can also be achieved by the addition of chloroform (20 ml per liter) of fermentation broth instead of NP40. Alternatively, cells may be collected by centrifugation prior to lysis, resuspended to 1 kg wet weight/6 l in Tris-EDTA as described in the first two methods and then lysed by the addition of NP40 or chloroform. Following cell lysis by either method the lysate is centrifuged in a continuous rotor at 26,000xg as described in the first two methods. As with those methods, LiBr extraction of the pellet and  $\text{NH}_4\text{SO}_4$  precipitation of the supernate are used to recover the product.

### 25 Purification of SlpIII

The pellet obtained by centrifugation of the cell lysate at 26,000xg as described above is extracted with an equal volume of 9M LiBr. The salt solution is added and the pellet is evenly suspended by stirring at room temperature (RT). The mixture is stirred for 1 hour at RT after an even suspension is obtained. The mixture is then centrifuged at 26,000xg in a continuous rotor at 4°C or at RT to generate a pellet and a supernatant fraction. The supernate is saved and the pellet is re-extracted with another equal volume of 9M LiBr as above. After mixing for 1 hour the mixture is centrifuged at 26,000xg and the supernate from this centrifu-



gation is combined with the supernate from the first LiBr extraction and allowed to stand at 4°C overnight. Approximately 90% of the SlpIII contained in the cell lysate 26,000xg pellet is extracted by LiBr using this procedure.

After the LiBr extract stands overnight at 4°C a precipitate forms, is removed by centrifugation at 26,000xg and is discarded. The supernate is then placed in dialysis bags and dialyzed against several changes of dH<sub>2</sub>O for 2 days. As the LiBr is removed by dialysis the SlpIII product precipitates in the dialysis bags. The precipitate is collected by centrifugation and washed 2-3 times with dH<sub>2</sub>O. The final washed product is centrifuged and dried by lyophilization.

For the recovery of SlpIII from the 26,000 g supernatant fractions, NH<sub>4</sub>SO<sub>4</sub> precipitation is used. Solid NH<sub>4</sub>SO<sub>4</sub> is slowly added to the sample which is maintained at 4°C, until 38% saturation is achieved (231 g/l). The mixture is then stirred at 4°C for 2-3 hours. The precipitate is recovered by centrifugation in a continuous flow centrifuge and washed 4-5 times with an equal volume of distilled H<sub>2</sub>O or with 0.5% SDS in H<sub>2</sub>O. After each wash the precipitate is recovered by continuous centrifugation. The pellet becomes increasingly white with successive washes as contaminating protein is removed. SlpIII is recovered as a washed pellet and can be dried by lyophilization.

#### 30      Trypsin Treatment Step of SlpIII

SlpIII was suspended in 50 mM Tris HCl, pH 8.0, 0.1 M NaCl buffer, and was placed in a 37°C water bath, and TPCK treated trypsin solution was mixed into the suspension. The final trypsin concentration was 0.1%. After 3 hours, the solution was centrifuged at 16,000xg for 15 min., the pellet was washed with a half equal volume of 0.5% SDS in H<sub>2</sub>O first, then with

distilled water. After each wash the pellet was recovered by centrifugation. The final product was resuspended in water and kept at 4°C for further analysis.

- 5           With the trypsin treatment, SlpIII was purified to 99.4% purity.

#### Physical Measurements of SlpIII

- Physical measurements of the purified silk-like proteins have been compared with those of Bombyx mori silk in order to establish that the repetitive amino acid polymers produced microbiologically accurately mimic the properties of naturally occurring polymers. Physical measurements were performed to confirm the model of anti-parallel chain pleated sheet conformation for the crystalline regions of Bombyx mori silk fibroin (Marsh, Corey and Pauling, Biochem. Biophys. Acta (1955) 16; Pauling and Corey, Proc. Natl. Acad. Sci. USA (1953) 39:247. Preliminary analysis of x-ray diffraction patterns obtained from Slp films are consistent with those described by Fraser, MacRae, and Steward (1966) (Table 4). Circular Dichroic (CD) and Fourier transform infrared (FTIR) spectroscopic analysis of SlpIII are consistent with a high degree of extended  $\beta$  and  $\beta$ -turn conformations. Comparisons of the spectra obtained from SlpIII with that of naturally occurring silk fibroin in various solvents (Isuka and Young, Proc. Natl. Acad. Sci. USA (1966) 55:1175) indicate that SlpIII in solution consists of a mixture of the random and highly ordered structures seen in silk fibroins.
- 10  
15  
20  
25  
30

Table 4

| <u>Material</u>       | <u>a (Å)</u> | <u>b (Å)</u> | <u>c (Å)</u> |
|-----------------------|--------------|--------------|--------------|
| (AG) <sub>n</sub>     | 9.42         | 6.95         | 8.87         |
| (ACAGSG) <sub>n</sub> | 9.39         | 6.85         | 9.05         |
| CTP fraction          | 9.38         | 6.87         | 9.13         |
| Native fibroin        | 9.40         | 6.97         | 9.20         |
|                       | 9.44         | 6.95         | 9.30         |
| SlpIII                | 9.38         | 6.94         | 8.97         |

Referenced in Fraser et al., J. Mol. Biol. (1966)  
19:580.

Example 4EBSI Gene Construction:

Six oligonucleotide strands were synthesized  
 and purified as described previously.

- i.  $\begin{matrix} \text{(HindIII)} & \text{BamII} & \text{SmaI} \\ 5' \text{AGCTGGCGCTCTGGAGTACGGCTGG} 3' \end{matrix}$
- ii.  $\begin{matrix} 5' \text{AATTCAGCGCTACTCCAGAGCGG} 3' \\ \text{(EcoRI)} & \text{SmaI} & \text{BamII} \end{matrix}$
- iii.  $\begin{matrix} \text{(HindIII)} & \text{BamI} \\ 5' \text{AGCTTGGTGGCAGGTGTAGGAGTTCCGGGTGTAGGCGTTCCGGGAGTTGG} \\ \text{TGTACCTGGAGTGGGTGTTCAGGCGTAGGTTGTCC} 3' \end{matrix}$
- iv.  $\begin{matrix} \text{(XbaI)} \\ 5' \text{CGCGGGCACACCTACGCGTGGAAACACCCACTCCAGGTACACCAACTCCCGGA} \\ \text{ACGCGTACACCGGAACTCCTACACCTGGGACCTAG} 3' \\ \text{BamI} \end{matrix}$
- v.  $\begin{matrix} \text{(XbaI)} & \text{AhaII} \\ 5' \text{CGCGGGGTAGGAGTACCAGGGGTACGGCTCCCTGGAGCCGGTCTCTGGTAG} \\ \text{CGCGGCAGGCGCGCGCTCCGGAGTACGGGTGGCGG} 3' \\ \text{BamII} & \text{BamI} \end{matrix}$
- vi.  $\begin{matrix} \text{(EcoRI)} & \text{BamI} & \text{BamII} \\ 5' \text{AATTCGGGACCCCTACTCCGGAGCGCGCGCGCTACCGACACCGG} \\ \text{CTCCAGGGACCGCTACCCCTGGTACTCCTACCG} 3' \\ \text{AhaII} \end{matrix}$

Oligonucleotide strands (iii), (iv), (v) and (vi) were annealed and ligated with the DNA of plasmid pBSm13(+) (Stratagene) which had been digested with HindIII and EcoRI. The products of this ligation reaction were transformed into E. coli strain JM109. Transformant colonies were selected for resistance to ampicillin. Colonies were screened for their hybridization with <sup>32</sup>P-labelled oligonucleotides (iii), (v). Plasmid DNA from several positively hybridizing clones was purified and sequenced. Two of the plasmids, pSY1292 and pSY1293, contained the sequence shown for oligonucleotides (iii), (v) and (iv), (vi). These sequences contained all of the nucleotides present in this synthetic oligonucleotides except one. A G:C basepair was missing at position 7 (iii). The lack of this basepair obstructed one of the BanI sites. In order to introduce a second BanI site at the 5' end of the gene fragment, oligonucleotides (i) and (ii) were annealed and ligated with plasmid pBSm13(+) which had been digested with HindIII and EcoRI. Plasmid DNA from the transformant colonies resistant to ampicillin was purified. Two plasmids, pSY1295 and pSY1296, which were digestible with StuI, a unique site contained in the oligonucleotide sequence, were sequenced. They were both shown to contain the sequence shown for oligonucleotides (i) and (ii). Plasmid DNA from pSY1292 was digested sequentially with HindIII, SI nuclease, and EcoRI. The digestion products were separated by electrophoresis in an agarose gel and the DNA fragment of approximately 150 basepairs was excised from the gel. This DNA fragment was digested with plasmid DNA pSY1296 which had been digested with StuI and EcoRI. The products of this ligation reaction were transformed into E. coli strain JM109 and were selected for resistance to ampicillin. Colonies were screened

for hybridization to  $^{32}$ P-labelled oligonucleotide (v). The plasmid DNA from two positively hybridizing clones was purified and sequenced. These plasmids were named pSY1297 and pSY1298. They contained the following

5 sequence:

(HindIII) BanII  
 AGCTGGGCTCTGGAGTAGGTGTCCAGGTGTAGGAGTTCCGGGTGTAGGCCTTCCGGGAG 60  
 TCGACCCGAGACCTCATCCACACGGTCCACATCCTCAAGGCCACATCCGCAAGGCCCTC

10

KnaI  
 TTGGTGTACCTGGAGTGGGTGTTCACAGGCGTAGGTGTGCCGGGGTAGGAGTACCAGGGG 120  
 AACCACATGGACCTCACCACAAAGGTCCGCATCCACACGGGGCCCATCCTCATGGTCCCC

15

BanII  
 TAGGCCTCCCTGGAGCGGGTGTGGTAGCGGCGCAGGCGCGGGCTTCCGAGTAGGGGTGC 130  
 ATCCGACGGGACCTCGCCACGACCATCGCCGCTCCGCGCCCGAGGGCTCATCCCCACG

20

EcoRI  
 CGAATTC  
 GCTTAAG

### EBSI Multimer Gene Assembly:

The BanI acceptor plasmid pSY937 was modified  
 25 in order to accept BanII terminal cohesive DNA frag-  
 ments. Two oligonucleotides were synthesized for this  
 purpose.

30

(BamHI) DraI SspI NruI BanII  
 vii 5'GATCCTATGTTTAAATATTCGCGAACGTTTTGTATCGGCTCGATGTGT  
 TACCGTCCGCGATCGATATCAGCTG3'  
 FspI EcoRV PvuII

35

(BamHI) PvuII EcoRV FspI BanII  
 viii 5'GATCCAGCTGATATCCATCGCGACCGTAACACATCGAGCCCATACAAAAA  
 CGTTCGCGAGAAATATTTAAACATAG3'  
 NruI SspI DraI

Oligonucleotides (vii) and (viii) were annealed and ligated with plasmid DNA pSY937 which was digested with BamHI. The products of this ligation were transformed into E. coli strain JM109 and colonies were selected for resistance to chloramphenicol. Transformant colonies were screened by hybridization to <sup>32</sup>P-labelled oligonucleotide (vii). Plasmid DNA from two positively hybridizing clones, pSY1299 and pSY1300, contained the sequence shown for oligonucleotides (vii) and (viii), as determined by DNA sequencing.

Plasmid DNA pSY1298 was digested with BanII and the digestion fragments separated by agarose gel electrophoresis. The EBSI gene fragment, approximately 150 base pairs, was excised and purified by electroelution and ethanol precipitation. Approximately 1 µg of purified fragment was self-ligated in order to produce multimers ranging in size from 450 bp to 6,000 bp. The products of the self-ligation were then ligated with plasmid DNA pSY1299 which had been digested with BanII. The products of this ligation reaction were transformed into E. coli strain HB101. Transformants were selected for resistance to chloramphenicol. Plasmid DNA from individual transformants was purified and analyzed for increased size due to EBSI multimer DNA insertions. Ten clones (pSY1240-1249) with inserts ranging in size from 1.5 Kbp to 4.4 Kbp were obtained.

### 30 Expression of EBSI Multimer Gene:

One of these clones, pSY1248, which contained a 4 Kb EBSI multimer gene was recloned in the λP<sub>R</sub> expression vector, pSY751. Plasmid DNA from pSY1248 was digested with NruI and PvuII, separated by agarose gel electrophoresis, and the DNA band corresponding to the EBSI multimer gene was excised and purified by NACS purification. DNA from plasmid pSY751 was digested

with PvuII and ligated with the NruI-PvuII fragment from pSY1248. The products of this ligation were transformed into E. coli HB101, and the transformants selected for resistance to ampicillin. Two clones were isolated containing the new plasmid pSY1280. E. coli cells containing pSY1280 were grown at 30°C to an OD<sub>600</sub> of 0.7 and then shifted to 42°C for 1.5 hours. The proteins produced by these cells was analyzed by SDS-PAGE. The separated proteins were transferred to nitrocellulose paper and detected by immunoreactivity with anti-ELP rabbit serum. A strongly reactive protein band was observed with an apparent molecular weight of 120 kD.

The Ampicillin drug resistance gene of pSY1280 was substituted with the Kanamycin marker and the subsequent plasmid was called pSY1332. This plasmid was used in fermentation for the purification of EBSI. (See Methods)

|    |                 |              |         |            |
|----|-----------------|--------------|---------|------------|
| 20 | pSY1332/pSY1280 | EBSI Protein | 1413 AA | MW 113,159 |
|----|-----------------|--------------|---------|------------|

MDPVVLQRRDWENPGVTQLNRLAAHPPFASERFCMGS

[(GVCVP)<sub>8</sub> (GAGAGSGAGAGS)<sub>1</sub>]<sub>26</sub>

MCYRAHGYQLSAGRYHYQLVWCQK

25

Purification of EBSI Protein:

E. coli strain HB101 containing plasmid pSY1280 was fermented in 10L volume. The cells were concentrated by filtration and further harvested by centrifugation. Pelleted cells were stored frozen at -70°C until processed. Frozen cells were thawed on ice and suspended in 4 ml of 50 mM Tris-HCl pH 7.0, 10 mM EDTA, 5 mM PMSF per gram wet weight of cells. The cells were broken by French pressing twice at 15,000 psi and then cooled to 0°C. The crude lysate was cleared by centrifugation at 26Kxg for 20 minutes. The supernatant proteins were precipitated by addition of

solid ammonium sulfate to 20% of saturation (114 g/l). The precipitate was collected by centrifugation at 10Kxg for 10 minutes. The pellet was resuspended in 10 ml of H<sub>2</sub>O and dialyzed against 10 mM Tris pH 8.0, 0.15M NaCl at 4°C. The dialyzed solution was digested with 0.1% Trypsin (Sigma) for 1.5 hours at room temperature, and reprecipitated with 20% ammonium sulfate. The precipitated protein was resuspended in H<sub>2</sub>O and dialyzed against 10 mM Tris pH 7.0, 1 mM EDTA at 4°C. The protein purity of this sample was analyzed by amino acid composition and determined to be 83%.

#### Elastic Properties of EBSI Protein:

The soluble preparation of semi-purified EBSI protein described above was incubated at 37°C for 30 minutes and centrifuged at 10Kxg for 10 minutes at room temperature. This treatment caused the EBSI protein to aggregate, become insoluble, and pellet into a translucent solid. The solid was resistant to mechanical disruption either by vortexing or by maceration using a glass rod. The solid could be cut with a razor blade into strips which exhibited a high degree of elasticity. They fully retained their shape after repeated extensions and relaxations. They resisted compression with no apparent irreversible deformation of structure.

#### EBSI Purification

EBSI sample (~70% pure) was dialyzed in 50 mM Tris HCl, 50 mM NaCl, pH 8.0 at 4°C overnight with one change of buffer. If precipitation was observed, the sample was centrifuged at 27,000xg for 15 min at 4°C. All remaining steps were performed at 4°C. The supernatant was applied to a DEAE-Sephacel column which had been equilibrated with 50 mM Tris HCl, 50 mM NaCl, pH 8.0. The flow through fractions which contained EBSI were collected and pooled. NaCl was added to the



pooled fractions from DEAE-Sephacel column to make a final concentration of 2 M NaCl in the sample.

Insoluble material was removed by centrifugation at 27,000xg for 20 min. The supernatant was then loaded

5 onto Phenyl-Sepharose column which was equilibrated with 50 mM sodium phosphate buffer, pH 7.0, with 2 M

NaCl. The column was washed extensively with buffer until no eluting protein was detected by A<sub>280</sub>. The

column was then eluted stepwise with 50 mM sodium

10 phosphate buffer, pH 7.0 and finally with water. The

EBSI active fractions were pooled and stored at 4°C for further analysis.

With the addition of these steps to the previous procedures, 100% pure EBSI was obtained.

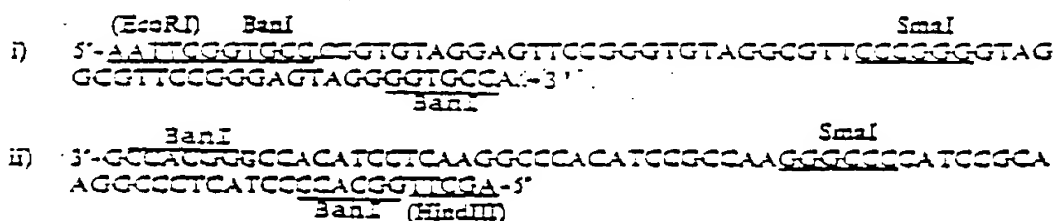
15

#### Example 5

#### ELPI Construction and Expression

Two oligonucleotide strands were synthesized and purified as described in the Methods section.

20



25

30

35

The two oligonucleotide strands were annealed and ligated with the DNA of plasmid pBS m13(+) (Stratagene) which had been digested with RENs HindIII and EcoRI.

5 The products of this ligation reaction were transformed into E. coli strain JM109. Transformant colonies were screened for their hybridization with  $^{32}\text{P}$ -labeled oligonucleotide (i). Plasmid DNA from positively hybridizing clones was purified and  
10 sequenced. One plasmid, pSY1287, contained the sequence shown for oligonucleotides (i) and (ii).

Plasmid DNA from pSY1287 was digested with BanI REN and the digestion fragments were separated by agarose gel electrophoresis. The ELPI gene fragment,  
15 approximately 60 bp, was excised and purified by NACS column. Approximately 1  $\mu\text{g}$  of purified fragment was self-ligated in order to produce multimers ranging in size from 300 bp to 5000 bp.

The products of the self-ligation were then  
20 ligated with plasmid DNA pSY937 which had been digested with REN BanI. The product of this ligation reaction was transformed into E. coli strain HB101. Transformants were selected for resistance to chloramphenicol. Plasmid DNA from individual transformants was purified  
25 and analyzed for increased size due to ELPI multiple DNA insertions. Four clones (pSY1388-1391) with inserts ranging in size from 1.0 kbp to 2.5 kbp were obtained. These clones were recloned in the  $\lambda\text{P}_\text{R}$  expression vector pSY751. The clones obtained  
30 (pSY1392-1395) were used for expression of ELPI.

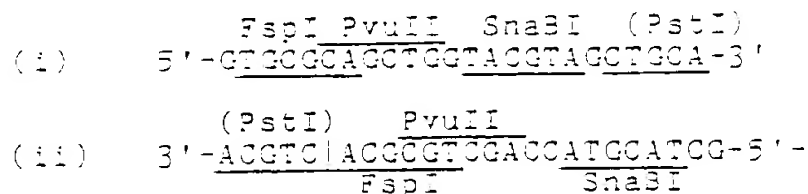
The ELPI protein had the following amino acid composition:

pSY1395 ELPI Protein 859 AA MW 72,555

MDPVVLQRRDWENPGVTQLNRLAAHPPFARNILAIRW  
[(VPGVG)<sub>4</sub>]<sub>40</sub> VPWTRVDLSAGRYHYQLVWCQK

SELP1 Gene Construction and Expression

Two oligonucleotide strands were synthesized and purified as described in the Methods section.



These oligonucleotide strands were annealed and ligated with plasmid pSY1304 which had been digested with PstI REN (pSY1304 differs from pSY857 by having a monomeric unit in place of the trimeric unit of pSY857). Plasmid DNA from transformant colonies resistant to chloramphenicol was purified. One plasmid, pSY1365, which was digestible with REN SnaBI, was sequenced and proven to be correct.

ELPI gene fragment purified as described (ELPI construction and expression) was treated with Mung Bean Nuclease as described by supplier (Stratagene). The DNA fragments mixture was then ligated with plasmid DNA pSY1365 which had been digested sequentially with RENs FspI, SnaBI and calf intestinal phosphatase. The products of this ligation reaction were transformed into E. coli strain HB101 and were selected for resistance to chloramphenicol. Plasmid DNA from individual transformants was purified and analyzed for the ELPI monomer DNA insertion. Two plasmids, pSY1366 A and B, were sequenced. They were both shown to contain the ELPI DNA sequence in the correct orientation.

Plasmid DNA pSY1366 was digested with REN BanI and the DNA fragment containing the SELP1 monomer was gel purified. To create multimers, 1 µg of the SELP1 DNA fragment was self-ligated. Multimers were obtained ranging in size from 500 bp to 10 kbp. The SELP1 multimers were cloned into the BanI site of pSY1262.

Positive clones were characterized by gel electrophoresis for the size of the inserted multimer and used for expression and protein analysis.

5

pSY1396 SELP1 Protein 2025 AA MW 148,212

MDPVVLQRRDWENPGVTQLNRLAAHPPFASDPMGAGS (GAGAGS)<sub>6</sub>

[GAA(VPGVG)<sub>4</sub> VAAGY (GAGAGS)<sub>9</sub>]<sub>24</sub>

10 GAA(VPGVG)<sub>4</sub> VAAGY (GAGAGS)<sub>2</sub> GACAMDPGRYQLSACRYHYQLVWCQK

#### SELP2 - Monomer Construction

Plasmid DNA pSY1298 was digested with BanII REN and the EBSI gene fragment was purified as described previously. The EBSI monomer fragment was ligated into pSY1304 (pSY937 containing a monomer of SlpIII, constructed as pSY857) which had been digested with BanII REN and treated with calf intestinal phosphatase).

20 The products of the ligation mixture were transformed in E. coli strain HB101. Transformants were selected for resistance to chloramphenicol. After restriction analysis of several isolates, one plasmid was chosen, pSY1301, containing a DNA fragment  
25 corresponding to the EBSI monomer gene.

#### SELP2 - Multiple Gene Assembly and Expression

Plasmid DNA pSY1301 was digested with REN BanI and the DNA fragment containing the SELP2 "monomer" was gel purified. To create multimers, 1 µg of the SELP2  
30 DNA fragment was self-ligated. Multimers were obtained greater than 12 kb in size.

The SELP2 multimers were cloned into the BanI site of pSY1262. Positive clones were characterized by  
35 gel electrophoresis for the size of the inserted

multimer. The clones with inserts ranging in size from 1.5 kb to 11 kb were selected. Plasmid DNA pSY1372 containing an insert of 6 kb (18 repeats) was used for further analysis and protein purification.

5

#### SELP2 - Protein Purification

E. coli strain HB101 containing plasmid pSY1372 was fermented according to the procedure described in Methods for fermentation. The cells were  
10 harvested by centrifugation. Pelleted cells were stored frozen at -70°C until processed. Frozen cells were thawed on ice and suspended in 4 ml of 50 mM Tris-HCl, pH 7.0, 10 mM EDTA, 5 mM PMSF per gram wet weight of cells. The cells were broken by passing through a  
15 Gaulin cell disrupter at 8,000 psi. The crude lysate was cleared by centrifugation at 26,000xg for 20 min. The supernatant, which contained >75% of the SELP2 protein, was precipitated by addition of 20% ammonium sulfate (114 g/L). The precipitate was collected by  
20 centrifugation at 10,000xg for 10 min. The pellet was resuspended in 10 ml of H<sub>2</sub>O and dialyzed against 10 mM Tris pH 8.0, 0.15 M NaCl at 4°C. The dialyzed material was centrifuged at 26,000x for 15 min in order to collect the insoluble fraction of protein which  
25 contained approximately 10% of the SELP2 protein. This insoluble protein pellet was washed twice in 0.2% SDS at 50°C for 30 min with occasional shaking. The insoluble protein was collected each time by centrifugation at 26,000xg for 15 min. followed by a wash of  
30 50% ethanol. The final protein pellet was resuspended in water and analyzed by Western blot analysis and amino acid composition. By Western blot the SELP2 protein appears to be homogeneous in size consistent with its large molecular weight (>150 kd). By amino  
35 acid composition the SELP2 preparation is approximately 80% pure and the observed molar ratio of amino acids

(Ser:Gly:Ala:Pro:Val:Tyr) agrees very closely with the expected composition as predicted from the SELP2 sequence present in pSY1372.

5 pSY1372 SELP2 Protein 2055 AA MW152,354

MDPVVLQRRDWENPCVTQLNRLAAHPPFASDPMCAGS (GAGAGS)<sub>2</sub> (GVCVP)<sub>3</sub>  
 [(GAGAGS)<sub>6</sub> GAAGY (GAGAGS)<sub>5</sub> (GVCVP)<sub>8</sub>]<sub>17</sub>  
 (GAGAGS)<sub>6</sub> GAAGY (GAGAGS)<sub>2</sub> GAGAMDPCRYQLSAGRYHYQLVWCQK

10

### SELP3 - Construction and Expression

Plasmid DNA pSY1301 was partially digested with REN HaeII and the digestion fragments separated by agarose gel electrophoresis. The larger DNA fragments  
 15 were excised and purified by NACS column. The purified fragments were self-ligated, the ligation reaction was heated at 70°C for 15' to inactivate the T4 DNA ligase and eventually digested with REN PstI. The digestion mixture was then transformed into E. coli strain JM109.  
 20 Transformants were selected for resistance to chloramphenicol. Plasmid DNA from individual transformants was purified and analyzed for: (1) resistance to REN PstI; and (2) deletion of 60 bp HaeII fragment contained within the SELP2 gene fragment. One clone  
 25 (pSY1377) satisfied both requirements. Plasmid DNA pSY1377 was digested with REN BanI and the DNA fragment containing the SELP3 monomer was gel purified. To create multimers, 1 µg of the SELP3 DNA fragment was self-ligated. Multimers were obtained ranging in size  
 30 from 500 bp to 10 kbp. The SELP3 multimers were cloned into the BanI site of pSY1262. Positive clones were characterized by gel electrophoresis for the size of the inserted multimer and used for expression and protein analysis.

35

pSY11397 SELP3 Protein 2257 AA MW 168,535

MDPVVLQRRDWENPGVTQLNRLAAHPPFASDPMGAGS (GAGAGS)<sub>2</sub>

[(GVGVVP)<sub>8</sub> (GAGAGS)<sub>3</sub>]<sub>24</sub>

5 (GVGVVP)<sub>3</sub> (GAGAGS)<sub>5</sub> GAGAMDPGRYQLSAGRYHYQLVWCQK

#### SLP4 - Construction and Expression

Plasmid DNA pSY1304 (pSY857 with a single monomeric unit as distinct from the trimeric unit of  
10 pSY857) was partially digested with REN HaeII and the digestion fragments separated by agarose gel electrophoresis. The larger DNA fragments were excised and purified by NACS column. The purified fragments were self-ligated, the ligation reaction was heated at  
15 70°C for 15' to inactivate the T4 DNA ligase and eventually digested with REN PstI. The digestion mixture was then transformed into E. coli strain JM109. Transformants were selected for resistance to chloramphenicol. Plasmid DNA from individual  
20 transformants was purified and analyzed for: (1) resistance to REN PstI; and (2) deletion of 60 bp HaeII fragment contained within the SELP2 gene fragment. One clone (pSY11378) satisfied both requirements. Plasmid DNA pSY11378 was digested with REN BanI and the DNA  
25 fragment containing the SLP4 monomer was gel purified. To create multimers, 1 µg of SLP4 DNA was self-ligated. Multimers were obtained ranging in size from 300 bp to 6 kbp. The SLP4 multimers were cloned into the BanI site of pSY11262. Positive clones were characterized by  
30 gel electrophoresis for the size of the inserted multimer and used for expression and protein analysis.

pSY11398 SLP4 Protein 1101 AA MW 76,231

35 MDPVVLQRRDWENPGVTQLNRLAAHPPFASDPMGAGS [(GAGAGS)<sub>6</sub>]<sub>27</sub>  
(GAGAGS)<sub>4</sub> GAGAMDPGRYQLSAGRYHYQLVWCQK

As is evident from the above results, highly repetitive sequences can be prepared, cloned, and used for expression to produce a wide variety of products which may mimic natural products, such as silk and  
5 other proteins and antigens. In addition, novel systems are provided for controlling the expression of the peptide under inducible conditions in a variety of hosts. In this manner, new proteinaceous products can be provided which provide for new properties or may  
10 closely mimic the properties of naturally occurring products.

### Bibliography

1. Maniatis, T., Fritsch, E.F. and Sambrook, J.  
15 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
2. Laemmli, U.K. 1970. Nature (London), 227:680-  
20 685.
3. Applied Biosystems User Bulletin. 1984. No. 13.
4. Matteucci, M.D. and Caruthers, M.H. 1981.  
25 Journal Amer. Chem. Soc., 103:3185-3319.
5. McBride, L.J. and Caruthers, M.H. 1983.  
Tetrahedron Letters, 24:245-248.
- 30 6. Smith, 1980. Methods in Enzymology, 65:371-379.
7. Vieira, J. and Messing, J. 1982. Gene, 19:259-268.
- 35 8. Anagnostopoulis, C. and Spizizen, J. 1981. J. Bacteriol., 317:41-746.



9. Davanloo, P., Rosenberg, A.H. Dunn, J.J. and Studier, F.W. 1984. Proc. Natl. Acad. Sci. USA, 81:2035-2039.
- 5 10. Rosenbluh, A., Banner, C.D.B., Losick, R. and Fitz-James, P.C. 1981. J. Bacteriol., 148:341-351.
11. Sadaie, Y., Burtis, K.C. and Doi, R. 1980. J. Bacteriol., 141:1173-1182.
- 10 12. Queen, C. 1983. J. Applied Molecular Genetics, 2:1-10.
- 15 13. Ferrari, F.A., Trach, K. and Hoch, J.A. 1985. J. Bacteriol., 161:556-562.
14. Johnson, W.C., Moran, C.P. and Losick, T.R. 1983. Nature (London), 302:800-804.
- 20 15. Studier, W.F. and Moffat, B.A. 1986. J. Mol. Biol., 189:113-130.
16. Goldfarb, D.S., Doi, R.H. and Rodriguez, R.L. 1981. Nature (London), 293:309-311.
- 25 17. Ferrari, F.A., Nguyen, A., Lang, D. and Hoch, J.A. 1983. J. Bacteriol., 154:1513-1515.
- 30 18. Lacey, R.W. and Chopra, I. 1974. J. Med. Microbiology, 7:285-297.
19. Norrander, J., Kempe, T. and Messing, J. 1983. Gene, 26:101-106
- 35 20. Sanger, F., Nicklen, S. and Coulson, A.R. 1977. Proc. Natl. Acad. Sci. USA, 74:5463-5467.

21. Biggin, M.D., Gibson, T.J. and Hong, G.F. 1983.  
Proc. Natl. Acad. Sci. USA, 80:3963-3965.

5 22. Zagursky, R.J., Baumeister, K., Lomax, N. and  
Berman, M.L. 1985. Gene Anal. Techn., 2:89-94.

23. Sanger, F. and Coulson, A.R. 1978. FEBS Letters,  
87:107-110.

10

24. Sadler, J.R., Techlenburg, M. and J. L. Betz.  
1980. Plasmids containing many tandem copies of a  
synthetic lactose operator. Gene 8:279-300.

15 All publications and patent applications  
mentioned in this specification are indicative of the  
level of skill of those skilled in the art to which  
this invention pertains. All publications and patent  
applications are herein incorporated by reference to  
20 the same extent as if each individual publication or  
patent application was specifically and individually  
indicated to be incorporated by reference.

The invention now being fully described, it  
will be apparent to one of ordinary skill in the art  
25 that many changes and modifications can be made thereto  
without departing from the spirit or scope of the  
appended claims.

30

35

WHAT IS CLAIMED IS:

1. A DNA sequence encoding a peptide containing an oligopeptide repeating unit, which repeating unit is characterized by containing at least three different amino acids and a total of from 4 to 30 amino acids, there being at least two repeating units in said peptide and at least two identical amino acids in each repeating unit and wherein said units are optionally joined by an amino acid bridge of from about 1 to 15 amino acids, said DNA sequence having the following formula:

$$K_k(WMX_XNY_Y)_iL_l$$

wherein:

K and L are each DNA sequences encoding an amino acid sequence of from about 1 to 100 amino acids, K and L being fewer than about 20 number % of the total amino acids;

k and l are 0 or 1;

W is of the formula:

$$[(A)_n (B)_p]_q$$

A is a DNA sequence coding for said oligopeptide repeating unit where A will contain from about 12 to 90nt, at least two codons coding for said identical amino acid in said repeating units being different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units may be the same or different;

n is an integer in the range of 1 to 100;

each  $p$  is independently 0 or 1; and  
 $q$  is at least 1 and is selected so as to  
 provide a DNA sequence of at least 90nt;

$M$  and  $N$  are the same or different and are a  
 5 DNA sequence of codons and are of 0 to 18nt in reading  
 frame with  $W$  and  $X$  respectively;

$X$  is the same as or different from  $W$  and is of  
 the formula:

$$10 \quad [(A^1)_{n1}(B^1)_{p1}]_{q1}$$

$Y$  is the same as or different from  $W$  and is of  
 the formula:

$$15 \quad [(A^2)_{n2}(B^2)_{p2}]_{q2}$$

wherein:

all of the symbols come within the definitions  
 of their letter counterparts;

20  $x$  and  $y$  are 0 or 1;

$i$  is 1 to 100; and

the total of  $q$ ,  $q^1$ , and  $q^2$  is at least 1 and  
 not greater than about 50.

25 2. A DNA sequence according to Claim 1,  
 wherein said DNA sequence is not greater than about  
 10knt.

30 3. A DNA sequence according to Claim 2,  
 wherein  $M$  and  $N$  are 0nt and at least one of  $x$  and  $y$  is  
 0.

4. A DNA sequence encoding a peptide  
 containing an oligopeptide repeating unit, which  
 35 repeating unit is characterized by containing at least  
 three different amino acids and a total of from 4 to 30  
 amino acids, there being at least two repeating units

in said peptide and at least two identical amino acids in each repeating unit and wherein said units are optionally joined by an amino acid bridge of from about 1 to 15 amino acids, said DNA sequence having the following formula:

$$K_k[(A)_n (B)_p]_q L_l$$

wherein:

10 K and L are each DNA sequences encoding an amino acid sequence of from about 1 to 100 amino acids, K and L being fewer than about 20 number % of the total amino acids;

k and l are 0 or 1;

15 A is a DNA sequence coding for said oligopeptide repeating unit where A will contain from about 12 to 90nt, at least two codons coding for said identical amino acid in said repeating units being different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units may be the same or different;

25 n is an integer in the range of 1 to 100; each p is independently 0 or 1; and q is at least 1 and is selected so as to provide a DNA sequence of at least 90nt.

30 5. A DNA sequence according to Claim 4, wherein said A codes for the same amino acid at least three times with at least two different codons.

6. A DNA sequence according to Claim 4,  
35 wherein said same amino acid is glycine.

7. A DNA sequence encoding a peptide containing an oligopeptide repeating unit, which repeating unit is characterized by comprising the amino acid sequence GAGAGS or GVCVP, said DNA sequence comprising the following formula:

$$[(A)_n(B)_p]_q$$

10 wherein:

A is a DNA sequence coding for said oligopeptide repeating unit, wherein the codons for at least two G's in the same or different A's are different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units may be the same or different;

20 n is an integer in the range of 5 to 25;  
each p is independently 0 or 1; and  
q is at least 1 and is selected so as to provide a DNA sequence of at least 90nt.

25 8. A DNA sequence according to Claim 7, wherein said oligopeptide is GAGAGS and B comprises a codon for tyrosine.

30 9. A DNA sequence according to Claim 7, wherein an A is of the formula:

GGTGGGGGGCAGGAAGT.

77

10. A DNA sequence according to Claim 8,  
wherein an A comprises the sequence:

GGT GCC GGC ACC GGT GCA GCA GCC GGT TCT GGA GCT GGC GCG  
5 GGC TCT GGC GCG GCG GCA GCA.

11. A DNA sequence according to Claim 7,  
wherein said oligopeptide is GVGVP.

10 12. A DNA sequence according to Claim 11,  
wherein one A is of the formula:

GCTGTAGGCCTTCGG.

15 13. A DNA sequence according to Claim 1,  
wherein WMX comprises the sequence:

GGA GTA GGT GTG CCA GGT GTA GGA GTT CCG GGT GTA GGC GTT  
CCG GGA GTT GGT GTA CCT GGA GTG GGT GTT CCA GGC GTA GGT  
20 GTG CCC GGG GTA GGA GTA CCA GGG GTA GGC GTC CCT GGA GCG  
GGT GCT GGT AGC GGC GCA GGC GCG GCG TCC.

14. A polypeptide comprising the recombinant  
DNA expression product of a DNA sequence encoding a  
25 peptide having repeating units of an oligopeptide,  
which oligopeptide is characterized by having at least  
three different amino acids and a total of from 4 to 30  
amino acids, there being at least two repeating units  
in said peptide and at least two identical amino acids  
30 in each repeating unit and wherein said units are  
optionally joined by an amino acid bridge of from about  
1 to 15 amino acids, said DNA sequence having the  
following formula:

$$K_k(WMX_xNY_y)_iL_1$$

35

wherein:

K and L are each DNA sequences encoding an amino acid sequence of from about 1 to 100 amino acids, K and L being fewer than about 20 number % of the total amino acids;

k and l are 0 or 1;

W is of the formula:

$$[(A)_n (B)_p]_q$$

10

A is a DNA sequence coding for said oligopeptide repeating unit where A will contain from about 12 to 90nt, at least two codons coding for said identical amino acids in said repeating unit being different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units may be the same or different;

n is an integer in the range of 1 to 100;

each p is independently 0 or 1; and

q is at least 1 and is selected so as to provide a DNA sequence of at least 90nt.

25

15. A polypeptide according to Claim 14, wherein said polypeptide will include as a repeating unit at least one of GAGAGS and GVGVP.

30

16. A peptide comprising the recombinant DNA expression product of a DNA sequence encoding a peptide having repeating units of an oligopeptide, which oligopeptide is characterized by having at least three different amino acids and a total of from 4 to 30 amino acids, there being at least two repeating units in said peptide and at least two identical amino acids in each repeating unit and wherein said units are optionally

35



joined by an amino acid bridge of from about 1 to 15 amino acids, said DNA sequence having the following formula:

$$5 \quad K_k [ (A)_n (B)_p ]_q L_l$$

wherein:

K and L are each DNA sequences encoding an amino acid sequence of from about 1 to 100 amino acids, K and L being fewer than about 20 number % of the total amino acids;

k and l are 0 or 1;

A is a DNA sequence coding for said oligopeptide repeating unit where A will contain from about 12 to 90nt, at least two codons coding for said identical amino acids in said repeating unit being different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units may be the same or different;

n is an integer in the range of 1 to 100;

each p is independently 0 or 1; and

q is at least 1 and is selected so as to provide a DNA sequence of at least 90nt.

17. A polypeptide according to Claim 16, wherein A codes for GAGAGS or GVGVP.

18. A regulatory system for regulated transcription in a prokaryotic host comprising:

(1) a DNA sequence which comprises a structural gene encoding an RNA polymerase exogenous to said host which is capable of transcribing DNA to messenger RNA and which is under the transcriptional control of an inducible promoter and

(2) a structural gene encoding a polypeptide of interest under the transcriptional control of a promoter which is not functional with the endogenous RNA polymerase of said host, but is  
 5 functional with said exogenous RNA polymerase.

19. A regulatory system according to Claim 18, wherein said inducible promoter is the beta-galactosidase promoter or the spoVG promoter.

10

20. A prokaryotic host comprising a DNA sequence according to Claim 1.

21. A method for producing a peptide  
 15 comprising the recombinant DNA expression product of a DNA sequence encoding a peptide having repeating units of an oligopeptide, which oligopeptide is characterized by having at least three different amino acids and a total of from 4 to 30 amino acids, there being at least  
 20 two repeating units in said peptide and at least two identical amino acids in each repeating unit and wherein said units are optionally joined by an amino acid bridge of from about 1 to 15 amino acids, said DNA sequence having the following formula:

25

$$K_k(WMX_xNY_y)_iL_l$$

wherein:

30 K and L are each DNA sequences encoding an amino acid sequence of from about 1 to 100 amino acids, K and L being fewer than about 20 number % of the total amino acids;

k and l are 0 or 1;

W is of the formula:

35

$$[(A)_n (B)_p]_q$$

A is a DNA sequence coding for said oligopeptide repeating unit where A will contain from about 12 to 90nt, at least two codons coding for said identical amino acids in said repeating unit being  
 5 different, where there will be at least two different A's differing by at least one nucleotide;

B is a DNA sequence different from A coding for other than the oligopeptide unit coded by the A unit and having from about 3 to 45nt, where the B units  
 10 may be the same or different;

n is an integer in the range of 1 to 100;  
 each p is independently 0 or 1; and  
 q is at least 1 and is selected so as to  
 provide a DNA sequence of at least 90nt;

15 M and N are the same or different and are a DNA sequence of codons and are of 0 to 18nt in reading frame with W and X respectively, encoding an amino acid sequence;

X is the same as or different from W and is of  
 20 the formula:

$$[(A^1)_{n^1}(B^1)_{p^1}]_{q^1}$$

Y is the same as or different from W and is of  
 25 the formula:

$$[(A^2)_{n^2}(B^2)_{p^2}]_{q^2}$$

wherein:

30 all of the symbols come within the definitions of their letter counterparts;

x and y are 0 or 1;

i is 1 to 100; and

the total of q, q<sup>1</sup>, and q<sup>2</sup> is at least 1 and  
 35 not greater than about 50;

said method comprising:

growing a prokaryotic host according to  
Claim 20,

wherein said DNA sequence is under the  
transcriptional and translational regulation of  
5 initiation and termination regulatory regions  
functional in said host, whereby said peptide is  
expressed; and

isolating the expression product.

10 22. A method according to Claim 21, wherein  
said peptide comprises at least one of the repeating  
units CAGAGS and GVGVP.

15

20

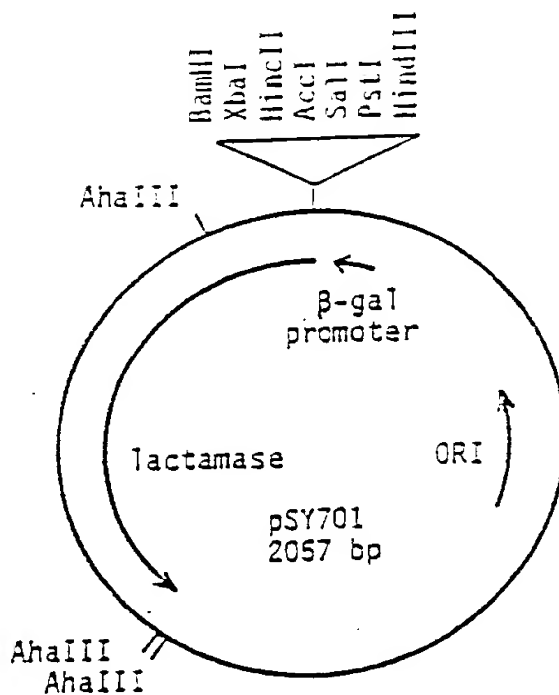
25

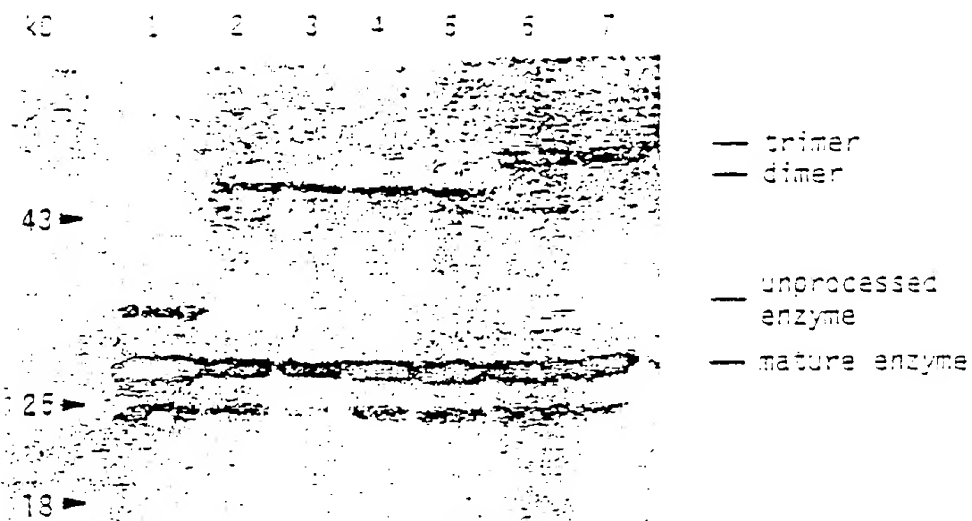
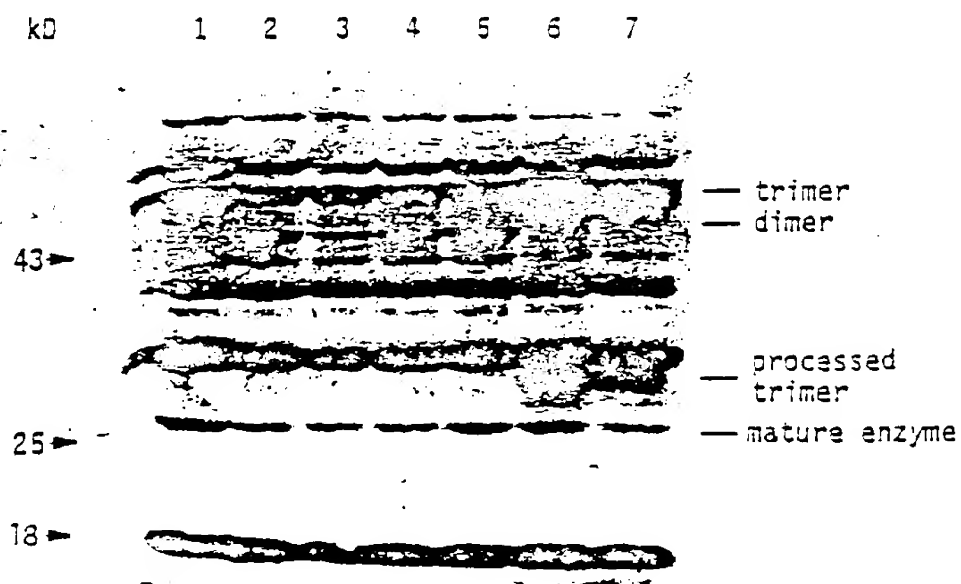
30

35

Figure 1

m t m i t p s l g c r s t l e d p h f r  
 ATGACCATGATTACGCCAAGCTTGGGCTGCAGGTCGACTCTAGAGGATCCCCATTTCGGT  
 : : : : :  
 HindIII PstI SalI XbaI BamHI  
 -1 +1 of mature  $\beta$ -lactamase  
 v a l i p f f a a f c i p v p a h  
 GTCGCCCTTATTCCCTTTTTCGCGGCATTTGCCTTCCTGTTTTTGCTCAC...



**A** Immunoblot with antibody to  $\beta$ -lactamase**B** Immunoblot with antibody to gly-ala peptide

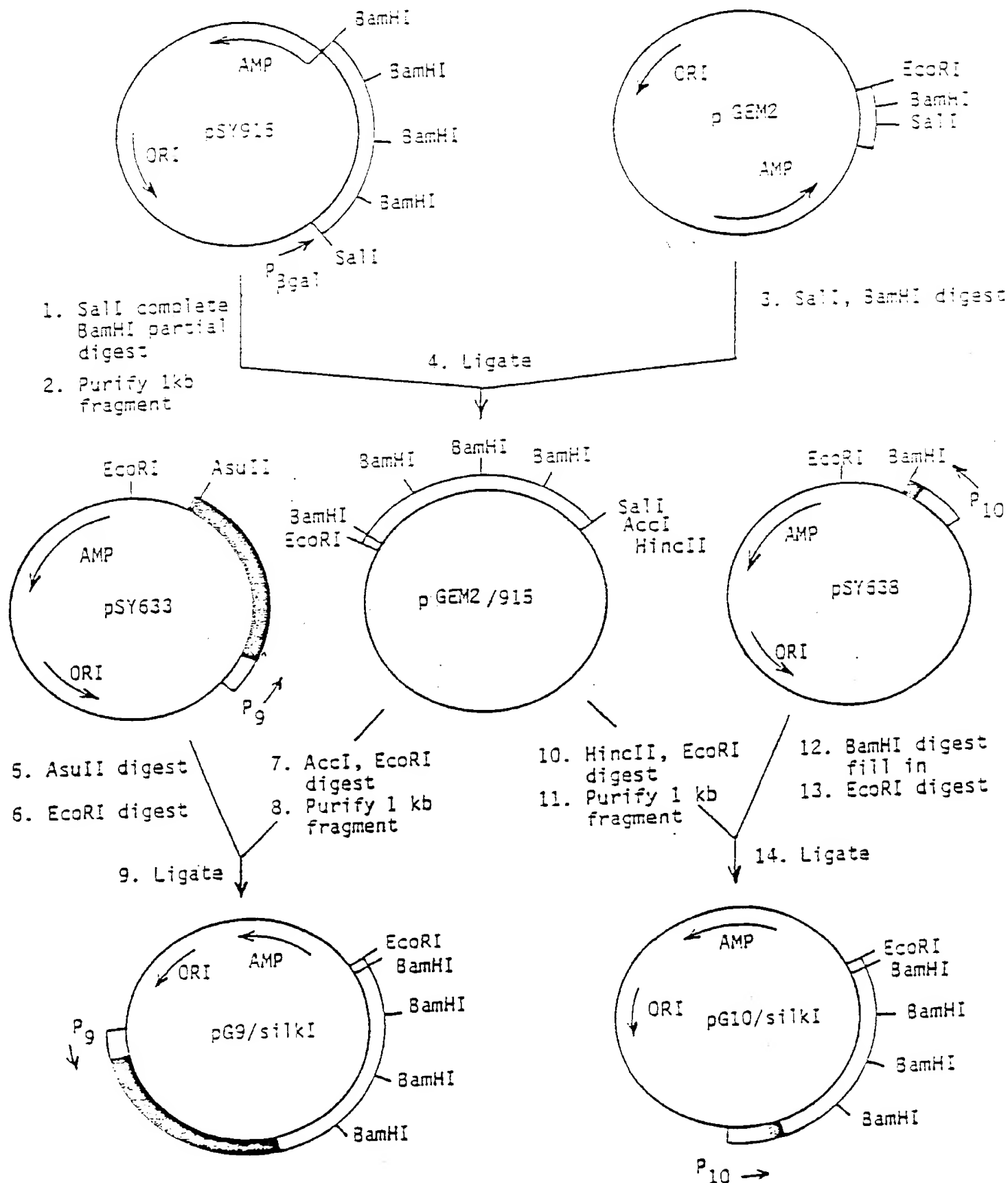


Figure 4

## A. T7gp10/silkI fusion: immunoblot with anti-silk Ab

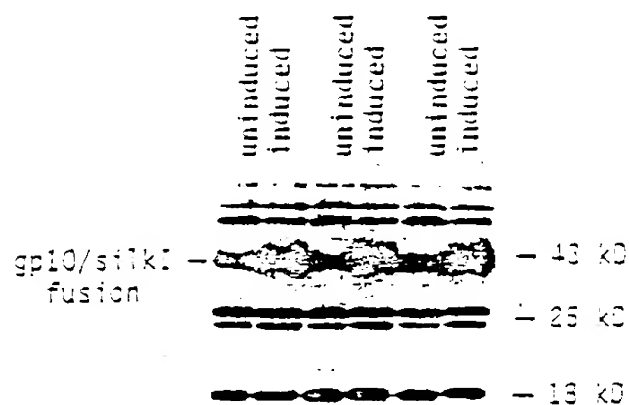
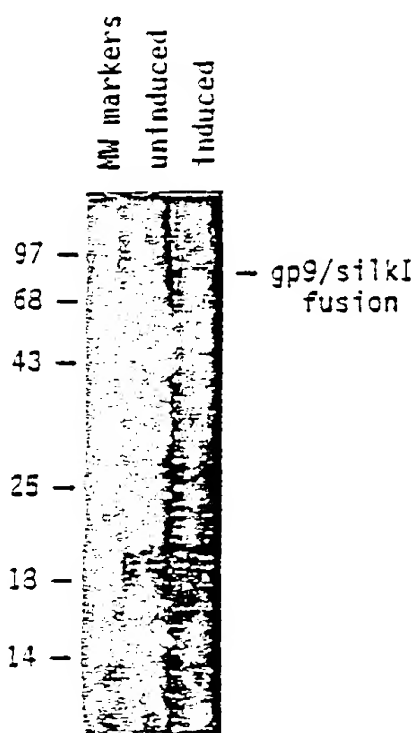
B. T7gp9/silkI fusion:  
immunoblot with anti-silk AbC. T7gp9/silkI fusion:  
Coomassie blue stain



Figure 5

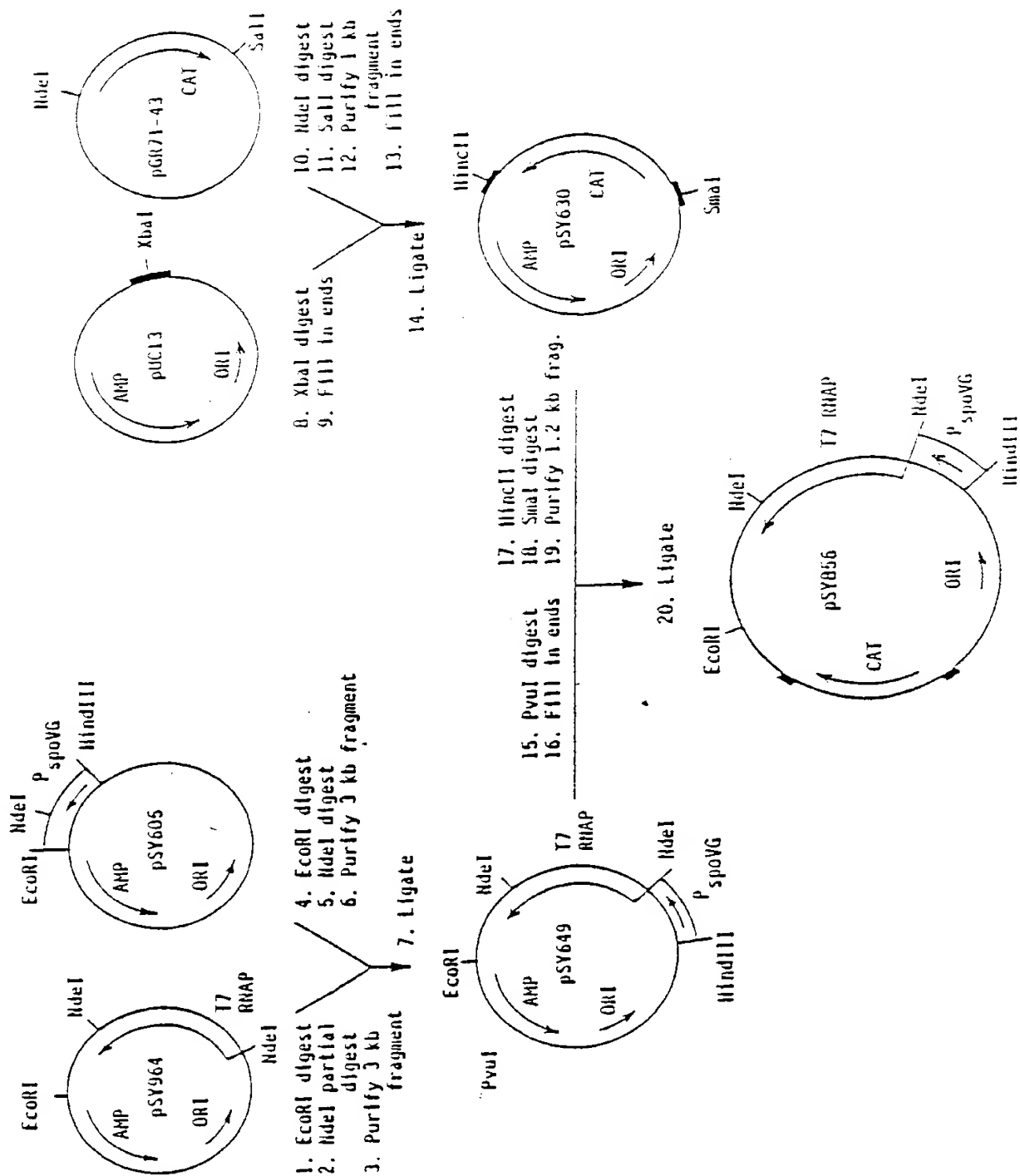


Figure 6

Timecourse for accumulation of the kanamycin-resistance gene product with the T7 system.

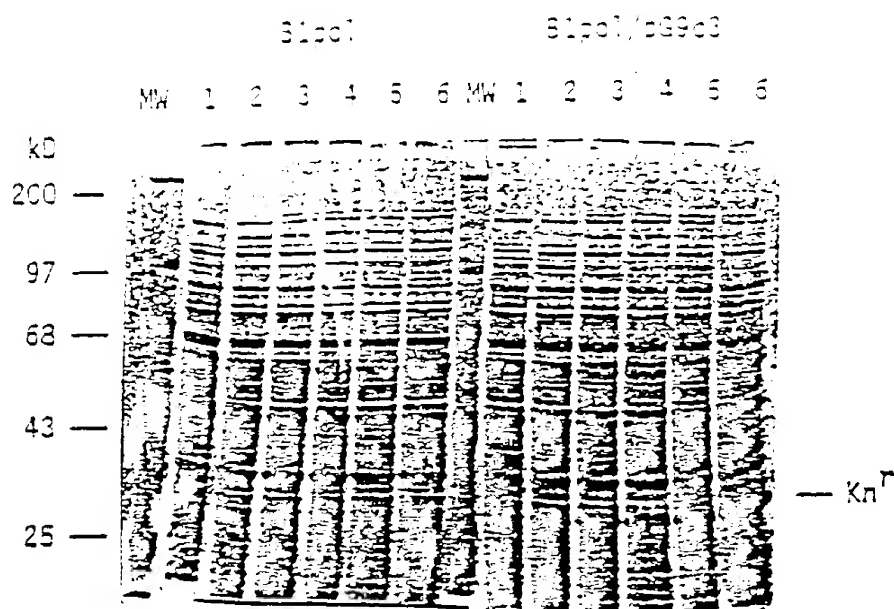
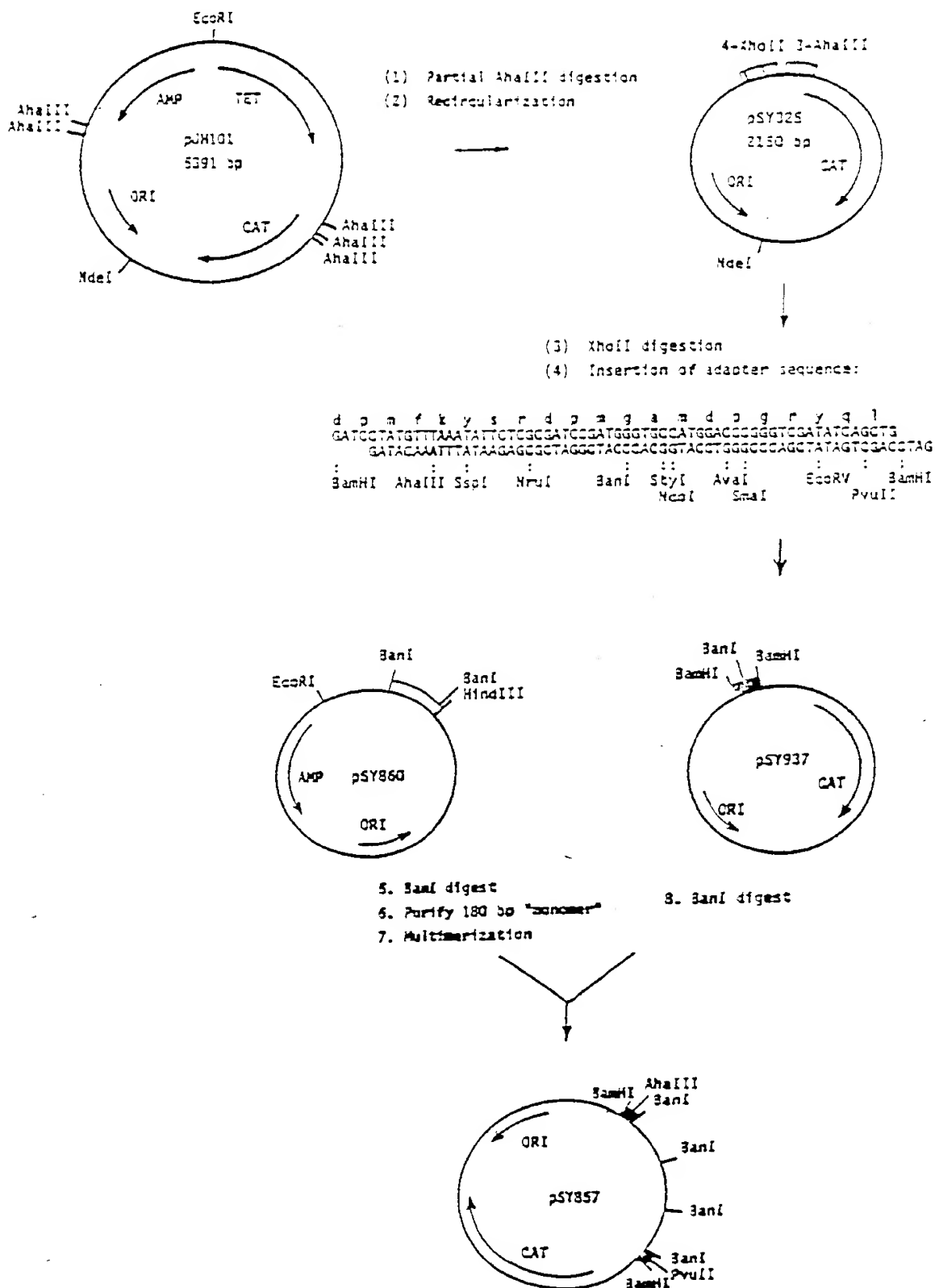


Figure 7



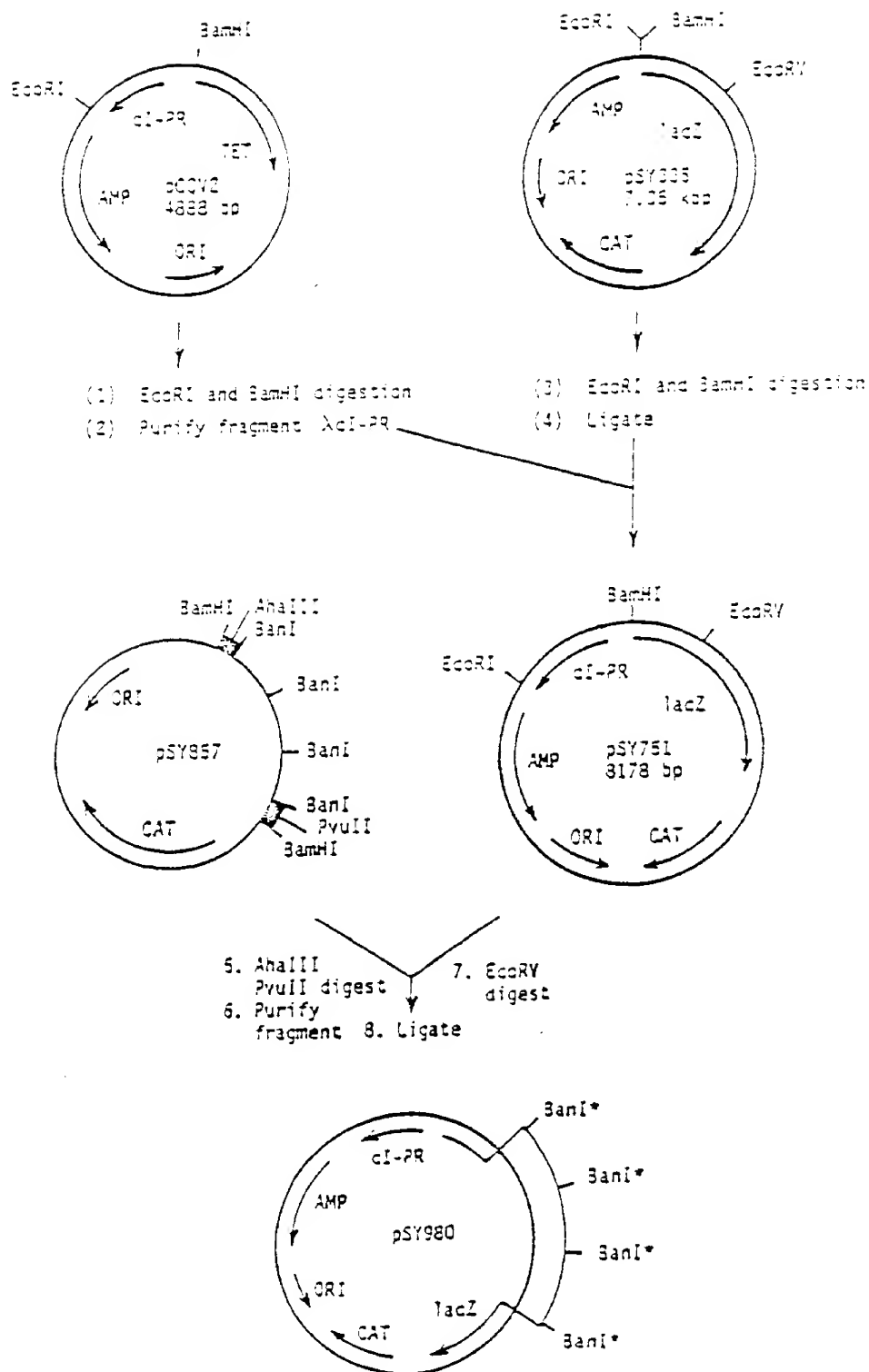
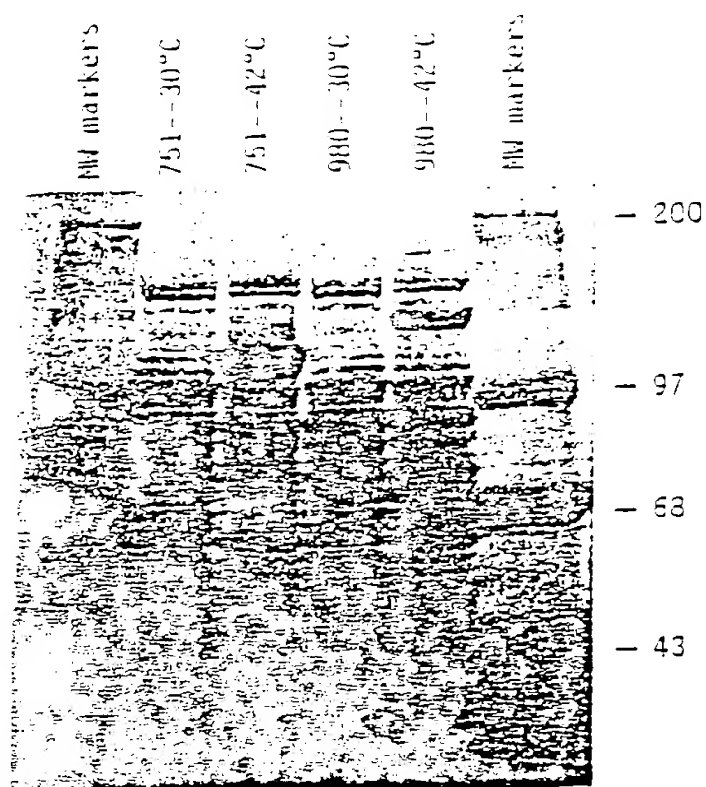
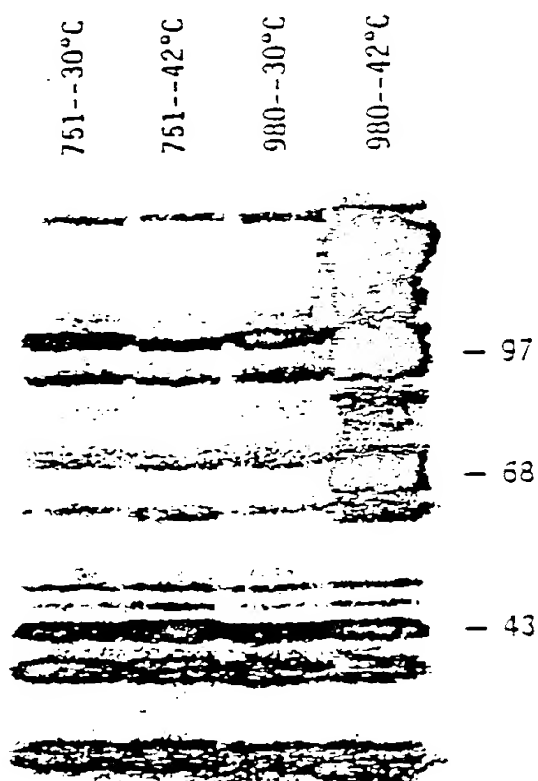


Figure 3

A.  $\beta$ -galactosidase/silkIII fusion: Amido black stainB.  $\beta$ -galactosidase/silkIII fusion: immunoblot with anti-silk Ab



# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US87/02822

|   |   |   |
|---|---|---|
| <b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) :<br>According to International Patent Classification (IPC) or to both National Classification and IPC<br>IPC(4) : C07H 21/04<br>US CL : 536/27  |   |   |
| <b>II. FIELDS SEARCHED</b>  |   |   |
| Minimum Documentation Searched *  |   |   |
| Classification System   | Classification Symbols  |   |
| U.S.  | 536/27, 28, 29; 435/68, 70, 71, 91, 172.1, 172.3, 320; 935/22, 29, 33, 36, 37, 39, 40, 41, 43, 61   |   |
| Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched *   |   |   |
| <b>CA SEARCH DATA BASE</b><br>CA FILE: 1967-88<br>BIOSIS FILE: 1967-88  |   |   |
| <b>III. DOCUMENTS CONSIDERED TO BE RELEVANT</b>   |   |   |
| Category *  | Citation of Document, <sup>1</sup> with indication, where appropriate, of the relevant passages <sup>2</sup>  | Relevant to Claim No. <sup>3</sup>                    |
| X<br>Y  | S. TABOR ET AL, "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes", Proceedings of the National Academy of Sciences USA, Volume 82, pages 1074-1078, published February 1985 by the National Academy of Sciences of the United States of America (Washington, D.C., USA). See entire document. | 18<br>19  |
| X   | F.W. STUDIER ET AL, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes", Journal of Molecular Biology, Volume 189, Number 1, pages 113-130, published 5 May 1986 by Academic Press (New York, NY, USA). See entire document.   | 18 and 19   |
| * Special categories of cited documents: <sup>4</sup><br>"A" document defining the general state of the art which is not considered to be of particular relevance<br>"E" earlier document but published on or after the international filing date<br>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)<br>"O" document referring to an oral disclosure, use, exhibition or other means<br>"P" document published prior to the international filing date but later than the priority date claimed<br>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention<br>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step<br>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art<br>"Z" document member of the same patent family |   |   |
| <b>IV. CERTIFICATION</b>  |   |   |
| Date of the Actual Completion of the International Search :   |   | Date of Mailing of this International Search Report : |
| 07 January 1988   |   | 02 FEB 1988   |
| International Searching Authority :   |   | Signature of Authorized Officer <sup>5</sup>          |
| ISA/US  |   | James Martinell                                       |

## FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

|   |  |    |
|---|--|----|
| Y | A. ROSENBLUH ET AL., "Identification of a new developmental locus in <i>Bacillus subtilis</i> by construction of a deletion mutation in a cloned gene under sporulation control", Journal of Bacteriology, Volume 148, Number 1, pages 341-351, published October 1981 by the American Society for Microbiology (Washington, D.C., USA). See abstract. | 19 |
|---|--|----|

☒ **OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE :**

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers ... because they relate to subject matter not required to be searched by this Authority, namely:

1-17 and 20-22  
 2. ☒ Claim numbers ... because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (2), specifically:

Claims 1-17 and 20-22 are unsearchable because they are so vague and indefinite as to be incomprehensible. Additionally, to the extent that the claims are understood, they embrace such a multitude of DNAs, proteins, and methods that no meaningful international search can be carried out. For example, to the extent understood, claim 1 embraces no fewer than  $1.2 \times 10^{216}$  different sequences.

☐ **OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING :**

This international Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.
2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:
3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

**Remark on Protest**

- ☐ The additional search fees were accompanied by applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.